Final Homework. [1]

Pankaj Mehta

May 6, 2019

The goal of this HW is to force you to think about all concepts we have learned in the class.

[1] This is the FInal Take-Home HW. **Due on Wednesday May 8th by 4pm. Please email me or leave in my office**. Please try to do this by yourself. You are allowed to discuss concepts with each other, you must write your own HW solutions independently. I will also be having office hours next week to help. I will be in my office on Monday May 6 from 9:30-1pm. for questions. You can also email me to set-up a time.

1. Growth laws revisited: Cell size regulation

We began the class discussing how we could make phenomenological growth laws. Here we will discuss simple models of cell-size control. The question is how does a cell maintain a constant volume. In other words, once there is a small fluctuation in the volume, why does it not grow? Why don't small changes in volume get amplified? This is now an active area of research (check out Willis and Huang, "Sizing up the bacterial cell cycle", Nature Reviews Microbiology 2017 for a good review).

In the absence of any regulation, smaller cells will have smaller daughter cells and larger cells will have larger daughter cells and cell sizes will diverge. Lets do a simple calculation and simulations to see how this works.Imagine a cell population with some average size $\bar{S}$. It is helpful to think about a general function F which relates the new cell size to old cell size

$$S_{i+1} = FS_i. \tag{1}$$

We can think of F as encoding complicated function that tells us about how sizes of mother and daughter cells are related.

If we denote the average population size by $\bar{S}$, it is actually useful to think about the linearized growth law since we are asking about how small perturbation grow of shrink:

$$\delta S_{i+1} = f\delta S_i, \tag{2}$$

where $S_i = \bar{S} + \delta S_i$. *(a)* Show that we can rewrite growth law equations as

$$S_{i+1} = fS_i + \bar{S}(1-f). \tag{3}$$

What are the different behaviors depending on the value of f? When is there cell-size regulation?

(b) Consider a population of $N = 100$ identical cells of size 1. Add a small perturbation to each cell drawn from a Gaussian distribution with mean zero and standard deviation 0.2. Simulate this equation for 50 generations for $f = 0, 1/2, 1, 2$. When running the simulation randomly choose a daughter cell to follow at each division (i.e. you

Figure 1: Figure of basic models of cell-cycle control.

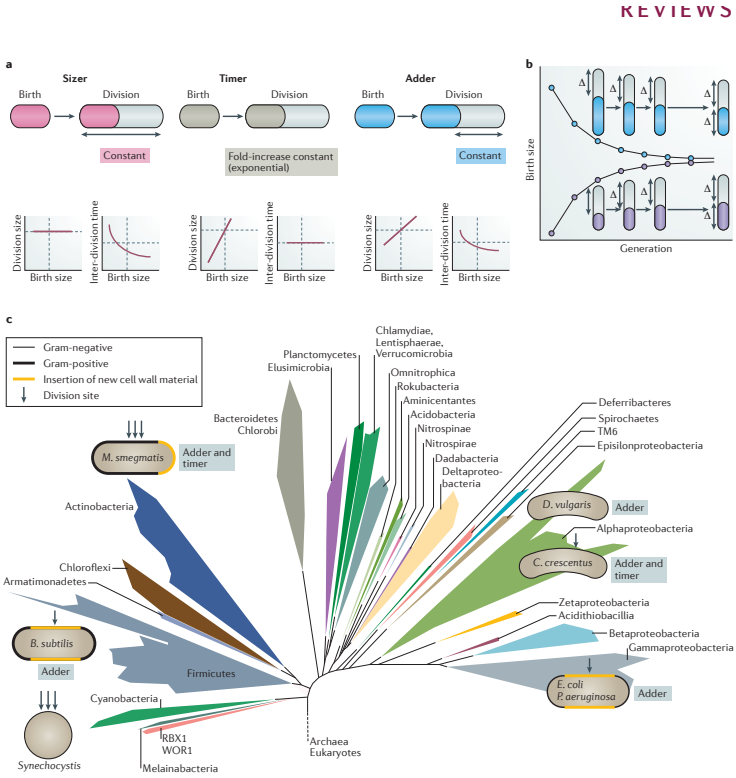

Figure 1 | **Quantifying the control of cell size.** Three simple models underlie cell size regulation (part **a**). In sizer regulation, cells divide (or trigger DNA replication) upon reaching a critical size. In timer regulation, cells divide after a fixed period after birth. In adder regulation, cells divide after the addition of a critical size to their birth size. Correlations between birth size and division size, as well as between birth size and inter-division time, can be used to distinguish among the three different models (red lines). Dashed lines denote the average values of the corresponding variables. The adder rule leads to progressive regression to the mean cell size over several generations (part **b**). Cells that are born larger than the average birth size (blue) add a fixed increment of material (Δ) that is equal to the average birth size and then divide. Thus, large newborn cells become smaller over several generations to approach the average cell size. Cells that are born smaller than the average birth size (purple) also add an increment (Δ) that is equal to the average birth size and then divide, so these cells increase their average size over several generations. Overlaid on a phylogenetic tree (part **c**) are the bacteria for which cell size control measurements have been reported to date. Cell outlines indicate Gram-positive (thick) or Gram-negative (thin) wall structure. Yellow regions indicate the places at which new cell wall material is inserted (if known). Arrows indicate the division sites. The reported homeostasis rule (or rules) is indicated. *B. subtilis*, *Bacillus subtilis*; *C. crescentus*, *Caulobacter crescentus*; *D. vulgaris*, *Desulfovibrio vulgaris* Hildenborough; *E. coli*, *Escherichia coli*; *M. smegmatis*, *Mycobacterium smegmatis*; *P. aeruginosa*, *Pseudomonas aeruginosa*. Part **c** is adapted with permission from REF. 155, Macmillan Publishers Limited.

should end up with 100 cells at the end of the simulation, with each cell a descendant of one of the 100 original cells). Make a plot of 10 of the cell lineages as a function of generation. Calculate the mean and standard deviation of the resulting distribution of cells? Explain your results.

(c) Decrease the standard deviation of the perturbation to 0.05 and redo *(b)*. How do your results change?

As shown in Figure **??**, there are there are three general models for how cells work. These are called the sizer (where cells directly control size), the timer (where cells grow for fixed time T), and a

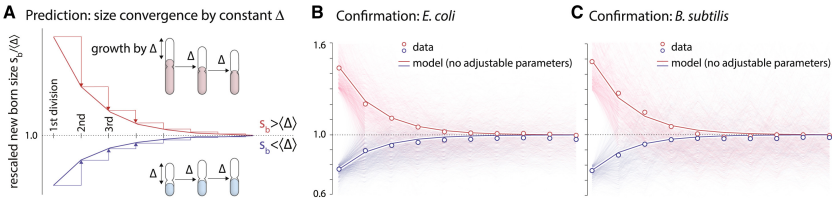

Figure 2: Real data confirming the adder model from Sattar Taheri-Araghi et al. Cell-Size Control and Homeostasis in Bacteria. Current Biology 2015.

adder (where cells add a constant volume at each cell division). Let us relate this to what we did above.

(d) Argue that the size corresponds to $f = 0$ and that the timer corresponds to $f = 1$.

(e) Let us now consider a simple "adder" model where one adds a fixed Δ to the cell at each iteration. In this case, we have

$$S_{i+1} = \frac{1}{2}(S_i + \Delta). \tag{4}$$

Show that this model corresponds to $\bar{S} = \Delta$ and that $f = 1/2$.

Note that we can check the type of mechanism used by a cell by just measuring f. This is actually much trickier than it seems and lot of fun experimental designs have been used. Here is data from Fig. **??** showing the adder describes *E. coli.* using a fun set-up they call the "Mother Machine".

2. **Random Walk model of Polymers** We spent a lot of time thinking about diffusion and random walks. It turns out that random walks are also good approximation for biological polymers like DNA, actin, and microtubules.

In particular, we can think of Polymers as being composed of N stiff rods of length l that perform a random walk (see Figure **??**). The reason we can do this is that each Polymer looks stiff on a short-enough scale. The length-scale associated with these stiff segments is usually called the Kuhn length. The simplest approximation to a polymer of length L is to think of it as $N = L/l$ stiff rods.

In what follows, for mathematical simplicity, we call consider a random walk in 2D dimensions. *(a)* Look up the Kuhn length for

DNA, actin, and microtubules in order to get a feeling for these polymers.

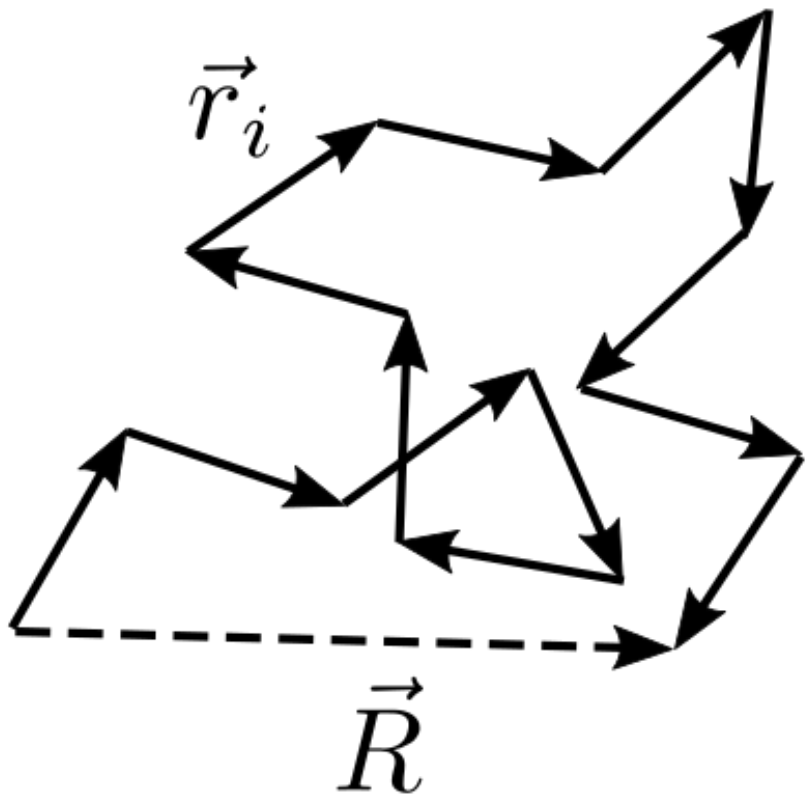

Figure 3: A polymer can be thought of as a random walk in three dimensions. This is called the Freely Jointed Chain. It of course ignores charge, steric occlusion, etc. but is still surprisingly useful model.

(b) Let us denote position by vector $\vec{R} = (R_x, R_y)$. Show that $\langle R_x \rangle = \langle R_y \rangle = 0$.

(c) Consider a single link $\vec{e} = (e_x, e_y)$ of length l (this is just an arrow of length l that can point uniformly in any direction on the 2d plane). Show that

$$\langle e_x^2 \rangle = \langle e_y^2 \rangle = \frac{l^2}{2} \tag{5}$$

(d) Since all links are independent, use this to calculate for a chain of N segments

$$\langle R_x^2 \rangle, \langle R_y^2 \rangle, \langle \vec{R}^2 \rangle. \tag{6}$$

(e) Use the derived formula to estimate the genome length (in μm) of *E. coli*.

3. **Molecular networks for calculating Berg-Purcell**

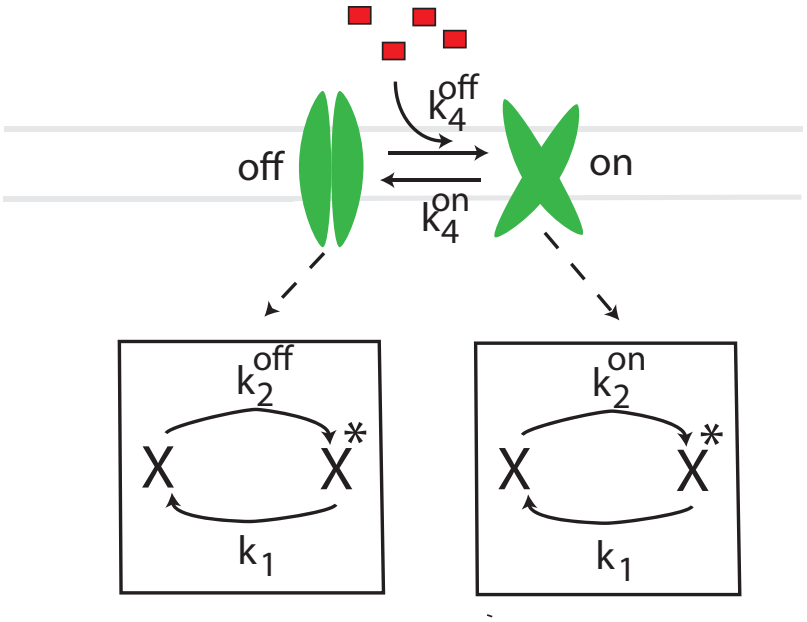

Figure 4: Figure of showing network that implments Berg-Purcell calculation

In this problem we will show that the simple two component circuit shown in Fig. **??** implements the Berg-Purcell algorithm for estimating concentrations we disscussed in class. This two-component biochemical network stores information about ligand concentration in the steady-state concentration of the activated form of a downstream protein (as shown in Figure **??**). Such two-component networks are a common signal transduction motif found in bacteria, and are often used to sense external signals through receptor-catalyzed phosphorylation of a downstream response regulator . The membrane-bound receptors can be in an either an active 'on' state or an inactive 'off' state. For simplicity, we assume that the binding affinity of the 'on' state is extremely high such that all ligand-bound receptors are always in the 'on' state and all unbound receptors are in the 'off' state. Receptors can switch between the off state and on state at a *concentration-dependent* rate k_4^{off} and from the on state to the off state at a concentration-independent rate k_4^{on}. Receptors additionally convert a downstream signaling protein from an inactive form X to an active form X^*, by for example phosphorylation, at a *state-dependent* rate k_2^s, where $s = $ on, off. The proteins are deactivated at a state-independent rate k_1. The dependence of k_2^s on the receptor state is what propagates information about ligand concentration from the receptor to X.

Importantly, the deactivation rate of the off state is small yet must be nonzero for thermodynamic consistency. We also note that for the case where proteins are activated through phosphorylaltion, k_2^{off}

includes non-specific phosphorylation arising from other kinases as well as contributions from the reverse reactions of the phosphatases. The inactivation rate sets the scale for the effective measurement time $T \propto k_1^{-1}$, since sets the rate at which information encoded in downstream proteins is erased due to inactivation. In order to compute external concentrations accurately, the measurement time must be much longer than the typical switching times between receptor states, $k_1 \ll k_4^{\text{on}}, k_4^{\text{off}}$. We show below that this simple network in fact implements a noisy version of the original Berg-Purcell calculation.

(a) Show that the equation describing the mean number of phosphorylated proteins takes the form

$$\frac{dp_{\text{on}}}{dt} = k_4^{\text{off}}(1 - p_{\text{on}}) - k_4^{\text{on}} p_{\text{on}} \tag{7}$$

$$\frac{dn}{dt} = k_2^{\text{on}} p_{\text{on}} + k_2^{off}(1 - p_{\text{on}}) - k_1 n. \tag{8}$$

(b) Show that at steady-state we have

$$\bar{p}_{\text{on}} = 1 - \bar{p}_{\text{off}} = \frac{K_4^{\text{off}}}{K_4^{\text{off}} + K_4^{\text{on}}} \tag{9}$$

and

$$\bar{n} = (K_2^{\text{on}} - K_2^{\text{off}})\bar{p}_{\text{on}} + K_2^{\text{off}}, \tag{10}$$

where we have defined the dimensionless parameters $K_j^s = k_j^s / k_1$ with $j = \{2, 4\}$ and $s = \{\text{'on','off'}\}$.

To quantify learning in our biochemical circuit, we follow Berg and Purcell and estimate the fluctuations in $(\delta c)^2$ as

$$\frac{\langle (\delta c)^2 \rangle}{c^2} = \left(c \frac{\partial \bar{n}}{\partial c} \right)^{-2} \langle (\delta n)^2 \rangle, \tag{11}$$

with $(\delta n)^2 = \langle n^2 \rangle - \bar{n}^2$.

(c) Let us assume that the on-rate is diffusion limited so that $k_4^{off} = k_+ c$ and denote $k_4^{on} = k_-$. Show that

$$\left(c \frac{\partial \bar{n}}{\partial c} \right)^2 = (\bar{p}_{on} \bar{p}_{off} \Delta K_2)^2. \tag{12}$$

Calculating $\langle (\delta n)^2 \rangle$ is actually quite tricky. However, one can do it using Generating Functions. Here, we will show the first few steps in the derivation.

(d) Show that the the master equation for the probability, $p_s(n)$, of there being n active proteins with the receptor in a state s,

$$\frac{dp_s(n)}{dt} = k_1(n+1)p_s(n+1) + k_2^s p_s(n-1) + k_4^{\bar{s}} p_{\bar{s}}(n) - (k_1 n + k_2^s + k_4^s)p_s(n) \tag{13}$$

where $\bar{s}$ = off (on) when s = on (off). State the meaning of each of these terms.

Let us define two generating function of the form

$$G_s(n) = \sum_{n=0}^{\infty} p_s(n) z^n, \tag{14}$$

with s = on, off.

(e) Show that the Master equation above can be re-written as

$$\left[(z-1)\partial_z - K_2^s(z-1) + K_4^s\right] G_s(z) = K_4^{\bar{s}} G_{\bar{s}}(z). \tag{15}$$

where we have defined $K_i^s = k_i^s/k_1$. It turns out that we can solve this differential equation using Hypergeometric equations and then solve for the variance. This is too hard for this HW [2]. Instead, we will state the answer.

[2] I encourage the more theory inclined graduate physics students to try to derive this expression below.

$$(\delta n)^2 = \bar{n} + \bar{p}_{on}\bar{p}_{off}\frac{(\Delta K_2^{on})^2}{1 + K_4^{on} + K_4^{off}} \tag{16}$$

(f) Write a Gillespie algorithm to check this answer for the parameters: $k_2^{\text{off}} = 0.01, k_2^{\text{on}}$, $k_4^{\text{on}} = k_4^{\text{off}} = k_1 = 1$.

Plugging this into the equation above gives

$$\frac{(\delta c)^2}{c^2} = \frac{\bar{n}}{(\bar{p}_{on}\bar{p}_{off}\Delta K_2)^2} + \frac{1}{(\bar{p}_{on}\bar{p}_{off})(1 + K_4^{on} + K_4^{off})}. \tag{17}$$

How do we relate this to Berg-Purcell? To compute uncertainty, Berg and Purcell assumed that the cell computes the average receptor occupancy by time-averaging over a measurement time T. They showed that,

$$\frac{(\delta c_{\text{BP}})^2}{c^2} = 2/N_b, \tag{18}$$

(g) Argue that number of binding events will equal the number of unbinding events and hence

$$N_b = T\bar{p}_{on}k_4^{\text{on}} = T\bar{p}_{off}k_4^{\text{off}} \tag{19}$$

(h) Let us identify the de-phosphorylation rate with the averaging time in BP as follows $k_1 = 2T^{-1}$. Show then that Eq. **??** can be written as

$$\frac{(\delta c)^2}{c^2} = \frac{\bar{n}}{(\bar{n} - K_2^{off})^2 p_{off}^2} + \frac{2}{N_b}\left(1 - \frac{k_1}{k_4^{on} + k_4^{off} + k_1}\right). \quad (20)$$

When the measurement time is much longer than the timescale of fluctuations in receptor activity, i.e. $k_4^{\text{on,off}} \gg k_1$ (or equivalently $K_2^{\text{on}} \gg K_2^{\text{off}} \gg 1$), and the average number of activated proteins is large, $\bar{n} \gg K_2^{off} \gg 1$, the expression above reduces to $(\delta c)^2/c^2 \approx 2/N_b$ in agreement with Eq. **??**.

4. **Monod-Wyman-Changeux (MWC) model of allosteric interactions**

This exercise introduces the Monod-Wyman-Changeux (MWC) model of allosteric interactions. The MWC model was first proposed to explain the sigmoidal response of hemoglobin to oxygen and has since become one of the canonical models allostery in biochemistry and biophysics. The main idea of the model is that an enzyme or protein can exist in multiple, interconvertible conformations with the probability that the enzyme is in a given confirmation determined by thermal equilibrium. The presence of ligands biases the enzyme towards one of these confirmations by shifting the relative free energies of the underlying protein confirmations.

In this exercise, we will derive the main results of the MWC model from simple thermodynamic and statistical mechanical arguments. This problem assumes knowledge of partition functions.

(a) Consider an protein with a single conformational state that can bind a ligand [L] from the environment. In thermal equilibrium, show that the free energy difference, ΔF, between the bound and unbound state is given by

$$\Delta F = -\log \frac{[\text{L}]}{K_D}, \quad (21)$$

with $K_D = k_-/k_+$, k_+ the ligand binding rate, and k_- is the ligand unbinding rate. K_D is called the binding affinity of the protein

(b) Now consider a protein that can exist in two states, an active state A, and an inactive state I. In the absence of ligand, the free energy of the active state is ϵ_A and the inactive state is ϵ_I. Furthermore, denote the binding affinity of the protein in the active state by K_D^A and the binding affinity in the inactive state K_D^I. Calculate the probability

that the protein is in the active state. Show that in the limit where ligand binding strongly favors the active state $K_D^I \gg [\mathrm{L}] \gg K_D^A$, this expression reduces to a form similar to the Michaelis Menten equation. Briefly discuss the meaning of K_M and the relationship to the Michaelis Metin equation.

(c) Generalize the calculation in **b)** to the case when the protein is composed on 2 independent, identical subunits each of which can bind ligand. For this case, there are 8 total possible states: the protein can be active or inactive with 0,1, or 2 ligands bound to the protein. Show that when $K_D^I \gg [\mathrm{L}] \gg K_D^A$, your expression reduces to a form similar to the Hill equation with a Hill coefficient of 2.