BE/APh161: Physical Biology of the Cell
Homework 2
Due Date: Wednesday, January 21, 2015

"Doubt is the father of creation." - Galileo Galilei

1. Number of mRNA

In this problem, we are going to work our way through an estimate of the number of mRNA molecules found in a bacterium and in a yeast cell. The idea of the estimate is to try to figure out over the entire set of genes in the organism, how many *total* copies of mRNA will be found in the cell. To do the estimate, we will first consider the case of a bacterium and then for yeast, we will make the assumption that things play out the same way and simply scale up our bacterial estimate. Our starting point is the number of proteins in a cell, which for a bacterium we take to be 3×10^6. This means that in order to make a new cell, this many proteins have to be synthesized in the 1000-3000 s of the cell cycle (depending upon growth conditions). If the ribosome translates at a rate of 20 aa/s, figure out a range of values for how many proteins each mRNA can crank out per minute. The range comes from how tightly packed the ribosomes are. What is the highest rate at which translation could occur (hint: think about the size of the ribosome and how tightly packed they can be)? Now use this to estimate the total number of mRNAs that are needed to supply the protein needed during a cell cycle. Provide estimates for both bacteria and budding yeast.

2. Diffusion

We have been interested in making estimates of the time scale associated with a number of biological processes. One of the most ubiquitous processes is diffusion, which is our "go to" null hypothesis for how molecules get around in cells.

In class I noted that the time scale for diffusing a distance L is given by $t = L^2/D$, where D is the diffusion constant. In this problem, we will formally derive this result. Note that parts (a) and (b) are effectively problem 13.2 of PBoC. Also, reading much of chap. 13 of PBoC will be very helpful

for doing this problem.

(a) Our goal is to find the diffusive profile for some molecular species as a function of time. If we are given an initial concentration, we can use the diffusion equation to determine the concentration distribution at a later time. To that end, consider the one-dimensional diffusion equation in free space given by

$$\frac{\partial c(x,t)}{\partial t} = D\frac{\partial^2 c(x,t)}{\partial x^2}. \tag{1}$$

In particular, consider that the initial concentration distribution is given by $c(x,0) = \delta(x)$, where $\delta(x)$ is the Dirac delta function and basically means that there is a spike at the origin. In particular, you will show that

$$G(x,t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}}, \tag{2}$$

where we introduce the Green function $G(x,t)$ to signify that this is the concentration profile for the special case in which the initial concentration is the spike at the origin as represented by the delta function.

To obtain the solution, we will Fourier transform the diffusion equation in the spatial variable x according to the Fourier transform convention

$$\tilde{f}(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)e^{-ikx}dx, \tag{3}$$

and

$$f(x) = \int_{-\infty}^{\infty} \tilde{f}(k)e^{ikx}dk. \tag{4}$$

Using these definitions, Fourier transform both sides of the diffusion equation to arrive at the ordinary differential equation

$$\frac{d\tilde{c}(k,t)}{dt} = -Dk^2\tilde{c}(k,t). \tag{5}$$

Solve this differential equation to obtain $\tilde{c}(k,t)$ and make sure to use the initial condition $c(x,0) = \delta(x)$ to find $\tilde{c}(k,0)$. Then invert the Fourier transform on $\tilde{c}(k,t)$ to find $c(x,t)$. NOTE: You will need to use completion of the square to carry out the inversion. Make sure you explain all of your steps. We are big on having you not only do the analysis correctly, but also to explain what you are doing and why you are doing it. Also, explain why

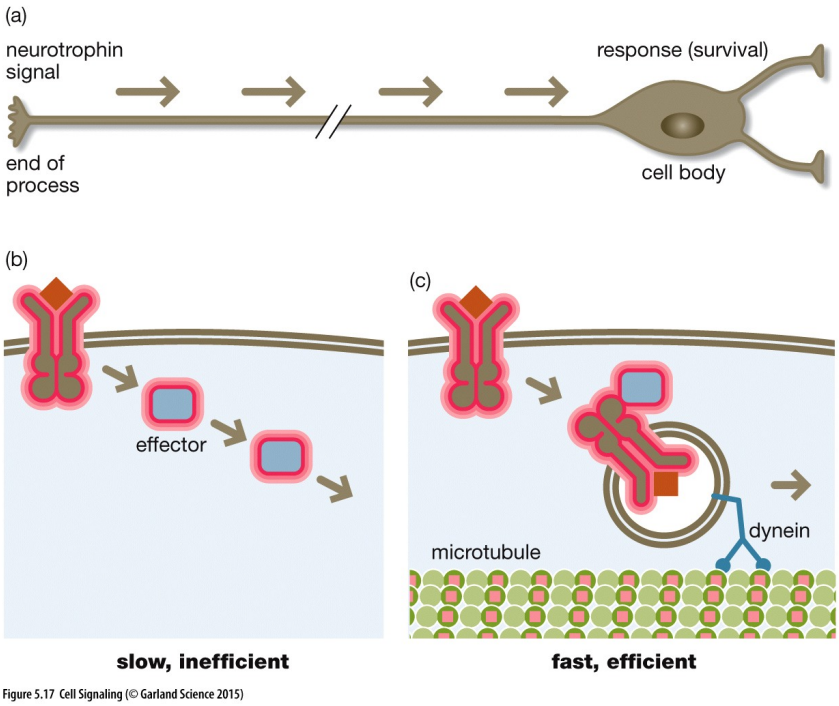

Figure 1: Comparison between passive diffusion and active transport in neurons. (a) Schematic of a neuron. (b) An effector molecule is activated and then diffuses along the axon to the cell body. (c) Receptor is incorporated into a vesicle and then actively transported by a dynein molecule along a microtubule.

I said this is the solution for "free space". Why would this solution fail to describe diffusion in a finite box?

(b) Using the solution we obtained above, find $\langle x \rangle$ and $\langle x^2 \rangle$. In general, we have that

$$\langle x^n \rangle = \frac{\int_{-\infty}^{\infty} x^n c(x,t) dx}{\int_{-\infty}^{\infty} c(x,t) dx}. \tag{6}$$

Explain what you find for both the first and second moments of the distribution as a function of time and explain how it relates to the estimated diffusion time $t = L^2/D$ which we use to find the time scale for diffusion over a length

L. Using the Einstein-Stokes relation given by

$$D = \frac{k_B T}{6\pi\eta a}, \tag{7}$$

where η is the viscosity which for water is $\eta_{water} = 10^{-3}\ Pa\ s$ and a is the radius of the diffusing particle, estimate the diffusion constant for a protein in water and make a log-log plot of diffusion time vs distance (with distances ranging from 1 nm to 1 m) and comment on its biological significance. Also, make a plot of the solution for the point source as a function of time by showing $c(x,t)$ at various times t using the same diffusion constant.

(c) In their book "Cell Signaling", Lim, Mayer and Pawson give the classic story about diffusion in neurons and how diffusion will take prohibitively long times. See Figure 1 for their depiction of the comparison between passive diffusion and active transport. Using what we have learned about diffusion, work out the time for diffusion of a protein over the 10 cm length of a neuron. Compare this to the time for a molecule to be transported actively by a motor. Do you agree with their assessment that active transport is efficient?

3. Flies by the numbers.

In this problem, like with our treatment of bacteria in class, we try to systematically explore some of the quantitative features of the *Drosophila* embryo.

(a) Make a sketch of an adult *Drosophila* with scale bars indicating the sizes of the head, wings and eyes. Using your sketch and your scale bars estimate the number of cells in the fly eye and the fly wing. For the eye, make sure you look at the structure of the eye and explain the key elements (see Figure 20.32 of PBOC2, for example). As usual, make sure you provide the rationale for your estimates.

(b) Do problem 20.2 of PBOC2. This part of the problem is intended to give a feeling for the time it takes to transcribe genes crucial for embryonic development.

4. Hill functions, the good, the bad and the ugly.

As discussed in class, a Hill function is of the form

$$p(x) = \frac{(\frac{x}{K_d})^n}{1 + (\frac{x}{K_d})^n}. \tag{8}$$

This function is used generically in the biological literature for a host of different processes where x is concentration and $p(x)$ could be the binding probability as a function of concentration, the activity of some molecule as a function of concentration or the probability that a ligand-gated ion channel is open as a function of concentration. Said differently, people are very indiscriminate in their uses of this function which ultimately makes it little more than an unsubstantiated fitting scheme.

(a) Plot such a function for the cases of $n = 1, 2$ and 4. Comment on what the "Hill coefficient" tunes.

(b) Imitating the argument for p_{bound} given in class and provided in Section 6.4.1 of PBOC2, consider a reaction involving a receptor with two binding sites. Imagine the reaction

$$\mathrm{L} + \mathrm{L} + \mathrm{R} \rightleftharpoons \mathrm{L_2R}, \tag{9}$$

where the notation L_2R means that the receptor is doubly bound. If we define the dissociation constant as

$$K_d^2 = \frac{[L]^2[R]}{[L_2R]}, \tag{10}$$

imitating the argument given in Section 6.4.1 of PBOC2, show that we find

$$p_{bound}([L]) = \frac{(\frac{[L]}{K_d})^2}{1 + (\frac{[L]}{K_d})^2}. \tag{11}$$

Interpret what it means to assume the chemical reaction in eqn. 9. Specifically, what does this whole procedure say about the states of single occupancy?

(c) Now let's redo the problem "correctly" by accounting for all of the states and their corresponding weights. What are the allowed states of this two-site receptor? Using our statistical mechanics approach described in class for simple ligand-receptor binding, work out the states and weights and find expressions for the probability of the empty state $p_0([L])$, the singly-occupied states $p_1([L])$ and the doubly occupied state $p_2([L])$. For the energy of the doubly occupied state, consider a total energy of the form $2\epsilon_b + \epsilon_{int}$, where ϵ_{int} is an "interaction energy" that imposes cooperativity in the binding. To do this problem and to make plots you will need to ascribe actual energies. Consider the case where $\epsilon_b = -5\ k_BT$ and $\epsilon_{int} = -2\ k_BT$. Make a single plot using Matlab that has p_0, p_1 and p_2. Use a log axis on the x-axis.

5. Dynamics of Populations.

In this problem, we are going to flex our muscles with the use of Matlab in order to look at the dynamics of populations. All of our estimates about the growth of *E. coli* have focused on the molecular processes that need to take place in order to grow cells. Interestingly, these molecular details can be ignored resulting in macroscopic growth equations such as

$$\frac{dN}{dt} = rN, \tag{12}$$

where $N(t)$ is the number of cells in our 5 mL growth tube (for example) as a function of time.

(a) Using a division time of 30 minutes, work out the value of the parameter r and then write a Matlab code and figure out the number of cells as a function of time given that in your 5 mL tube you started out with one cell at time $t = 0$. As an aside, given this growth rate, how long would it take for the mass of cells from repeated doubling to equal the mass of the Earth?

(b) Obviously, the model written above has the fatal flaw that in the long-time limit, it doesn't account in any way for the depletion of resources. To that end, we introduce the "logistic equation" for population growth that can be written as

$$\frac{dN}{dt} = rN(1 - \frac{N}{K}). \tag{13}$$

Like you did for the previous part of the problem, integrate this equation numerically in order to find the number of cells as a function of time. What

value of K should you use so that the number of cells saturates at 5×10^9?