BE/APh161: Physical Biology of the Cell
Homework 2
Due Date: Wednesday, January 30, 2013

"Every calling is a great calling when greatly pursued." - Oliver Wendell Holmes

RP to class: for all problems involving the use of some computer code, please submit your code with the homework.

1. Counting Proteins with Partitioning Statistics.

In this problem we consider the concentration of mRNA or proteins as a function of time in dividing cells. In particular, the point of this problem is to work out the concentration of mRNA or protein given that we start with a single parental cell that has N copies of this mRNA or protein (in the experiments of Golding *et al.* they watch the mRNA dilution effect while in the experiments of Rosenfeld *et al.* this is a fluorescently-labeled transcription factor). In the Rosenfeld experiment described in the paper you will read, at some point while the culture is growing, the production of the protein is stopped by providing a chemical in the medium and then the number of copies per cell is reduced as a result of dilution as the cells divide.

(a) Begin by reading the paper by Rosenfeld *et al.* entitled "Gene Regulation at the Single-Cell Level" (posted on the website with the homework) and write a one paragraph commentary on the paper with special reference to how they used the partitioning idea that is the subject of this paper. What is the experiment they did and what were they trying to learn?

(b) For this part of the problem, let's focus on the protein dilution effect. Work out a differential equation for the change in protein concentration as a function of the time that has elapsed since production of the protein was stopped. Solve the equation and make sure that your result depends upon the cell cycle time. Note that here we are only interested in the dilution that results from the original N copies of the protein being partitioned into an ever-larger number of daughter cells, not in the dilution that occurs as each individual cell lengthens in preparation for the next round of division. Note

also that in this part we're interested in a continuous model—you'll look at the discrete version in part (c).

(c) We can repeat a calculation like that given above using a discrete language. Imagine that before cell division, the number of copies of a given transcription factor in the cell is N. In particular, for every cell doubling, the number of proteins is reduced by a factor of 2. Using such a picture, write a formula for the average number of proteins per cell as a function of the number of cell divisions and relate this result to that obtained in part (b). Furthermore, by using the fact that $2 = \exp(\ln 2)$, reconcile the discrete and continuous pictures precisely.

(d) Interestingly, the model used in part (c) opens the door to one of the most important themes in physics, namely, that of fluctuations. In particular, as the cells divide from one generation to the next, each daughter does not really get $N/2$ copies of the protein since the dilution effect is a stochastic process. Rather the partitioning of the N proteins into daughter cells during division follows the binomial distribution. Analyzing these fluctuations can actually lead to a quantification of the number of copies of a protein in a cell. In this part of the problem, work out the expected fluctuations after each division by noting that the fluctuations can be written as $\sqrt{< (N_1 - N_2)^2 >}$, where N_1 and N_2 are the number of proteins that end up in daughter cells 1 and 2 respectively. Show that $\sqrt{< (N_1 - N_2)^2 >} = \sqrt{N}$ (hint: you'll need to use the binomial theorem.) I would like you to actually derive the result as was described in class rather than looking up the properties of binomial distributions. Also, remember that $N_1 + N_2 = N$.

(e) Next, look at the Rosenfeld paper and explain how measuring fluorescence variations can be used to calibrate the exact number of copies of the fluorescent protein in a cell. Assume that the fluorescence intensity in each cell can be written as $I = \alpha N$, where α is some calibration factor and N the number of proteins. Find a formula relating $\sqrt{< (I_1 - I_2)^2 >}$ and I_{tot}, where $I_{tot} = \alpha N$. Make a plot of $\sqrt{< (I_1 - I_2)^2 >}$ versus I_{tot} and explain how to get the calibration factor α from this plot.

(f) Extra credit: Now we are going to repeat the Rosenfeld experi-

ment numerically in order to *fit* the calibration factor. Consider a fluorescent protein such that the calibration factor between the intensity and the number of fluorophores is 50. Generate intensity data by choosing $N_1 + N_2 = 10, 50, 100, 1000$ and 5000 and for each case, "partition" the proteins from the mother cell to the two daughters 100 times (i.e. as if you are looking at 100 mother cells divide for each choice of the protein copy number). Then, make a plot of the resulting $\sqrt{< (I_1 - I_2)^2 >}$ vs I_{tot} just as we did analytically in the previous problem. What I mean is that you need to make a plot of all of your simulation results. Then, do a fit to your "data" and see how well you recover the calibration factor that you actually put in by hand. Plot the fit on the same graph as all of the "data".

2. Diffusive Time Scales

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 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.

3. Chemotaxis and Receptor Binding

As described in class, bacterial chemotaxis is claimed to be the best studied signal transduction problem in biology. In this problem, we work through some of the statements and results in a few of the classic papers I presented in class. We develop a feeling for the numbers by examining direct quotations from the experimental papers that have really driven the field recently as well as a commentary on this work by Dennis Bray. Begin by reading both of these papers which are attached on the website.

In their 2002 paper in PNAS entitled "Receptor sensitivity in bacterial chemotaxis", Sourjik and Berg say: "The changes in receptor occupancy encountered by bacteria swimming in spatial gradients (e.g., near the mouth of a capillary tube in the capillary assay) are very small. For example, in the tracking experiments, cells about 0.6 mm from the tip of a capillary tube (consider a pipette with a radius of 5.0 μm) containing 1 mM aspartate moved in a gradient of steepness 0.02 μM/μm at a mean concentration of about 8 μM. A 10-μm run straight up such a gradient would change the concentration from 8 to 8.2 μM, i.e., by 2.5 %. Assuming K_d values for aspartate of 7.1 μM and 62 mM (see above), this step gives a fractional change in receptor occupancy of about 0.003". **RP to class:** the two K_d values correspond to the fact that two of the different chemotactic receptors (Tar and Tsr) will bind aspartate, but with quite different affinities. For your estimates, only consider the smaller K_d since the larger one will be irrelevant at the concentrations of interest here.

Your job is to carry out calculations that exploit the numbers given above and using what you know about the definitions of concentration, the size of *E. coli* cells and about the meaning of K_d and simple binding curves (i.e. $p_{bound} = (L/K_d)/(1+(L/K_d))$, corresponding to the simplest model in which there is only a single binding site per receptor and there is no cooperativity between receptors). First, use the steady-state diffusion equation for a spherically symmetric source to estimate the concentration at 0.6 mm from

the pipette. The idea is to solve the 3D diffusion equation in spherical coordinates, given that the concentration at the source (i.e. the pipette) is 1 mM and that the concentration in the far field is zero. (NOTE: we work this out in chap. 13 of PBoC in a different context, but the ideas are all the same.) Do you agree with them about the concentration being 8 μM at a distance of 0.6 mm? Next, examine the statement about the consequences of a 10-μm run and also about the fractional change in occupancy. Do you agree with their numbers? Do you agree with the qualitative thrust of their statements?

In his commentary on the paper of Sourjik and Berg, Dennis Bray says: "The mystery can be expressed in a different way. Estimates of the binding affinity of aspartate to the membrane receptor of wild-type *E. coli* typically give a dissociation constant in the range 15 μM. A bacterium responding to a change in occupancy of 0.1% is therefore sensing concentrations of aspartate of a few nanomolar. And yet we know from decades of observations that the same bacterium is also capable of responding to gradients of aspartate that extend up to 1 mM. Somehow, *E. coli* is able to sense aspartate over a range of at least 5 orders of magnitude in concentration by using just one molecular species of receptor!"

Your job is to actually do the estimate/calculation that supports the claim made by Bray. In particular, examine a 0.1% change in occupancy and see what that means about the change in concentration given that the K_d has the value claimed. Also, if there is a change of concentration of order a few nanomolar, how many fewer molecules are there in a box of size 1 μm^3 due to such a concentration difference at the front and back of a cell?