BE/APh161: Physical Biology of the Cell
Homework 8
Due Date: Monday, March 12, 2012

"The wise are doubtful." - Socrates

1. Computing Transcriptional Decision Making.

Do problem 19.2(a), 19.3(a) and 19.3(b) from PBoC.

2. Counting Proteins with Partitioning Statistics.

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?

In this problem we consider the concentration of mRNA or proteins as a function of time in dividing cells. This exercise provides some of the conceptual tools we will need to write down rate equations describing gene expression. 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, 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) 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 (b). HINT: there are two ways to approach this problem. You can consider the change in the concentration as a function of the change in the number of cells into which the original N proteins are partitioned. Or you can note that for a bacterium like *E. coli*, it is a reasonable assumption to imagine that the cell diameter is unchanged and that the size is controlled by the cell length, such that the change in volume with time is simply the change in length with time times a constant prefactor; then consider the change in protein concentration as

a function of the change in the total volume into which the original N proteins are diluted.

(b) 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 (a). Furthermore, by using the fact that $2 = \exp(\ln 2)$, reconcile the discrete and continuous pictures precisely.

(c) Interestingly, the model used in part (b) 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.)

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. 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.

(d) Now we are going to repeat the Rosenfeld experiment 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".

3. Transcription and translation in development
(a) The average length of a gene in *Drosophila melanogaster* is about 11 kb and the average elongation rate of a transcript is about 1.2 kb/min. How does the time to produce an average mRNA compare to the nuclear cycle times in the

initial stages of fly development? For example, nuclear cycles 9 through 11 last no more than 6 minutes, while nuclear cycle 12 lasts about 10 minutes and cycle 13 on the order of 12 minutes. How do the genes that are actually expressed in this stage compare to the average gene in terms of their lengths? You can search for the size of these genes by going to `flybase.org` and searching for hb (*hunchback*), gt (*giant*), kr (*krüppel*) and kni (*knirps*).

(b) When *Drosophila* eggs are laid they already contain mRNA for several "maternal factors". Bicoid is an example of such a factor. Its mRNA is localized at the anterior end of the embryo, serving as a source of Bicoid protein. It is essentially stable up until the end of nuclear cycle 14 when it gets actively degraded. In this problem we want to estimate the number of mRNA molecules deposited in the embryo by its mother from measurements of the number of Bicoid proteins at nuclear cycle 14. Assume that all Bicoid is localized to the nuclei, which at cycle 14, approximately 120 minutes after the egg is laid, have a radius of about 3.3 μm. Use the fact that the bicoid gradient is of the form

$$c(x) = c_0 e^{-x/\lambda}, \tag{1}$$

with $c_0 = 60$ nM and $\lambda = 100$ μm to estimate the total number of Bicoid molecules in the whole embryo at this point. To get a sense of the size of the embryo and the number of nuclei, look at Fig. 2.35 in PBoC. Assuming that translation of *bicoid* mRNA is constant, estimate the number of mRNA molecules that led to your calculated number of Bicoid proteins. The number of ribosomes per kb on a transcript can be estimated by using the fact that there are about 13 ribosomes per Bcd mRNA. Bcd is 500 aa long.

4. Your Turn to Teach 161

RP: all answers to this problem must be submitted electronically to the TAs and me in pdf form.

Some have argued that only by quantitation will we really be able to come to terms with the complexity of living organisms, with the quantitative approach advocated in this class meant to give you a feel for how such quantitative dissection of biological problems might work. Others have argued that the approach we have taken is a mopping up operation which amounts to dotting the "i"s and crossing the "t"s already worked out by biologists. Write one paragraph defending each of these two points of view. One document you might find interesting to look at is "Bio2010" from the National Academy of Sciences.

Next, make a syllabus for the course. Start with one brief paragraph on the mission of your course. Issues that you might want to consider include: is it important to do hard calculations, or is that the province of other physics courses and our goal here is to illustrate the *style* of thinking? Are estimates a part of the way you will present the material (if yes, why, if no, why not?). How will you organize the material - note that in typical biology books DNA and actin

would never be in the same chapter but for PBoC they are both in chap. 10 as examples of "beam theory". The course is only 10 weeks long. What will you cover, what will you skip and why? How will you balance the desire to cover more topics with the resulting superficiality? This is not a look up something in Wikipedia question, nor is it a request to regurgitate what I did in the course. It is asking you how to organize a new and unfinished topic and to present it to advanced Caltech undergrads and to grad students at the beginning of their grad careers. What are the important points?