APh161: Physical Biology of the Cell
Homework 2
Due Date: Thursday, January 21, 2010

"Twenty years from now you will be more disappointed by the things you didn't do than by the ones you did do. So throw off the bowlines. Sail away from the safe harbor. Catch the tradewinds in your sails. Explore...dream...discover." Mark Twain

Comments from RP to Class:

This second homework continues with the theme of developing a feeling for the numbers in biology and supplements that with using the Boltzmann distribution and the physics of random walks to understand some biological processes.

Referee report: Read the second set of vignettes from the new book that I am writing with Ron Milo entitled "Cell Biology By the Numbers" that are posted on the course website associated with this homework and write a referee report on each one. Note that on the first homework some of you did an exquisite job while some small fraction of you gave reports that were at best cursory. I would encourage you to try something more thoughtful. Also, we have included a first cut at a "preface" that is intended to give you some sense of what the book is about. Read that as background for your reading of the remaining vignettes and provide any input you have on this first very casual attempt at a preface. As last time, the report should focus on the following questions: Does the overall logic make sense? That is, is the point of the vignette clear and does the organization work in making this point? What suggestions do you have to make it more readable, clear and interesting? Did it teach you anything new? What would you suggest should be removed? Try to find extra biological numbers pertinent to the vignette. Bonus: join the community effort and contribute these numbers at www.BioNumbers.org

Please E-mail the report, as a Word or PDF file, to me (phillips@pboc.caltech.edu), Stephanie Johnson (stephj@caltech.edu) and my coauthor Ron (ron.milo@weizmann.ac.il) on the day the homework is due.

1. Lost in Translation

In class I briefly touched on the idea of molecular recognition based strictly

on binding reactions. In this problem, you will work out a simple estimate using the Boltzmann distribution of the rate of errors in translation. Read section 18.4.1 of PBoC that considers biological fidelity. Consider the HP world in which there are only two kinds of amino acids, hydrophobic and polar, and explore the lowest error rates that can be achieved by simple thermodynamic binding reactions. Make sure you explain your simplified picture of the translation process in this world and how you are going to invoke equilibrium statistical mechanics to evaluate its fidelity. The essence of our analysis is the claim that the error rate is given by

$$\text{error rate} = \frac{r p_{err}}{r p_{err} + r p_{corr}}, \tag{1}$$

where r is the synthesis rate and p_{err} and p_{corr} are the probabilities of the binding reaction being the incorrect or the correct tRNA. Give a heuristic explanation of the equation given above - what assumptions are being made about the rates of binding and the rate of incorporation to arrive at this equation? Use the fact that the rate of addition of a given species is

$$\frac{dN_{err}}{dt} = r p_{err} \tag{2}$$

and

$$\frac{dN_{corr}}{dt} = r p_{corr}, \tag{3}$$

where N_{err} and N_{corr} are the number of incorrect and correct incorporations. As we did in chap. 18, take the binding energies to be ϵ_{err} and ϵ_{corr}, respectively. Using what you know about the translation process, make an estimate for $\Delta\epsilon = \epsilon_{corr} - \epsilon_{err}$ and then obtain the result that

$$\text{error rate} \approx \frac{p_{err}}{p_{corr}} \tag{4}$$

and make a plot of this error rate as a function of the energy difference. Make sure you explain why this definition of the error rate is the right idea. Further, explain why this result demonstrates that there has to be something more to molecular recognition in processes such as translation than mere thermodynamic discrimination. NOTE: In this problem, think of yourself as writing an explanation that could be read by someone else to learn about thermodynamic discrimination. Do not simply copy our way of doing it in the book and make sure you fill in missing steps and that you provide appropriate explanatory text.

2. Random Walks, Polymer "Size" and Genome Length.

The goal in this problem is to *estimate* the length of the *E. coli* genome in basepairs by examining an electron microscopy image of its exploded DNA. Chapter 8 of PBoC might be useful for carrying out these estimates and calculations.

(a) In class I made the claim that the size of the genome is given by

$$\text{size} \approx \sqrt{\langle R^2 \rangle} = a\sqrt{N}. \tag{5}$$

State all of the assumptions that go into deriving this expression and then derive it for yourselves. Explain both the merits and the problems with applying this kind of analysis to pictures like that of the *E. coli* genome given in fig. 8.6.

(b) As practice with the use of the binomial distribution and to make sure you feel the $\sqrt{N}$ result given above "in your bones", we will derive the end-to-end distance in one dimension by using the binomial distribution. Your goal is to compute the average end-to-end distance $\langle x \rangle$ and the standard deviation $\sqrt{\langle x^2 \rangle}$. Use the fact that $x = (n_r - n_l)a$ where n_r is the number of right pointing monomers and n_l is the number of left pointing monomers. Make a simple diagram of a one-dimensional random-walk configuration that shows the end-to-end distance and how it depends upon the number of monomers of each type. Explain your result for both of these averages and their significance for the estimate of the genome size.

HINT: Eliminate n_l by using the relation $N = n_l + n_r$. To proceed, show that the probability of having n_r steps pointing to the right out of a total of N steps is

$$p(n_r, N) = \frac{N!}{n_r!(N - n_r)!} p_r^{n_r} p_l^{n_l}, \tag{6}$$

where p_r is the probability of a right step and p_l is the probability of a left step. Note that later we will invoke $p_r = p_l = 1/2$, but for now leave this as is since it will help us with our analysis. Now, to obtain $\langle x \rangle$ and $\sqrt{\langle x^2 \rangle}$ all you need to do is compute $\langle n_r \rangle$ and $\langle n_r^2 \rangle$. Show that these averages are of the form

$$\langle n_r \rangle = \sum_{n_r=0}^{N} n_r \frac{N!}{n_r!(N - n_r)!} p_r^{n_r} p_l^{N-n_r}, \tag{7}$$

and

$$\langle n_r^2 \rangle = \sum_{n_r=0}^{N} n_r^2 \frac{N!}{n_r!(N-n_r)!} p_r^{n_r} p_l^{N-n_r}. \tag{8}$$

To evaluate these expressions, note that

$$\langle n_r \rangle = p_r \frac{\partial}{\partial p_r} \sum_{n_r=0}^{N} \frac{N!}{n_r!(N-n_r)!} p_r^{n_r} p_l^{N-n_r}. \tag{9}$$

But the binomial theorem tells us that

$$(p_r + p_l)^N = \sum_{n_r=0}^{N} \frac{N!}{n_r!(N-n_r)!} p_r^{n_r} p_l^{N-n_r}. \tag{10}$$

With these tips, you have everything you need to evaluate the averages you are asked to compute. When you get your result, make sure to connect it to the result of part (a) and to explain how this can be used to find the "size" of a one-dimensional DNA molecule.

(c) Using fig. 8.6, make a best estimate at the number of basepairs in the *E. coli* genome, remembering that the Kuhn length $a = 300$ bp. Compare your estimate to the observed genome length for *E. coli* and give a thoughtful discussion of how the estimate might have gone wrong.

3. Energy scales for photosynthesis and respiration.

In class, I made a big deal about the many interesting energy scales in biology. In this problem, you will explore two of these processes in more detail, again, with the objective just being to get a sense of the numbers so that you will carry them around with you for the remainder of the course (and hopefully after as well).

(a) Light is the giver of life. Work out the energy per photon for photons in the visible range. Make a plot showing the energy of these photons as a function of their wavelength. In addition, using the attached figure which shows the solar energy incident on Earth, estimate the number of photons with a wavelength of 800 nm arriving per square meter each second. Also, using this curve make an estimate of the power/m^2 for all of the incident photons.
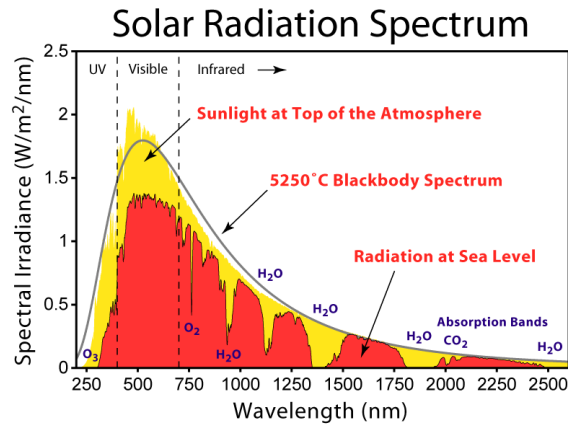
4

Figure 1: The solar spectrum. This figure shows the spectral density of light at the top of the atmosphere and on the surface of the Earth as a function of wavelength. (Adapted from `http://rredc.nrel.gov/solar/spectra/am1.5/` and `http://lasp.colorado.edu/ bagenal/3720/CLASS4/SolarSpectrum2.jpg`

(b) One of the outcomes of all of that busy photosynthesis is the synthesis of sugars which can then be used by organisms such as us in respiration. How much free energy is liberated in the combustion of a single glucose molecule? Report your answer in both Joules and $k_B T$ units. The cell has created biochemical pathways to harness the energy released from this reaction. By utilizing the glycolysis and oxidative phosphorylation pathways, cells can generate about 30 ATP molecules by converting one glucose molecule to carbon dioxide and water Let's examine the typical American diet that consists of 2000 kcal every day. Assuming this diet consists entirely of glucose, calculate how much ATP is synthesized every day. Give your answer in both moles and in kilograms. Finally, use this result to estimate how many molecules of ATP one cell in your body consumes each second.