MCB137L/237L: Physical Biology of the Cell
Spring 2020
Homework 3: A Feeling for the Numbers and Bacterial Growth
(Due 2/13/20 at 3:30pm)

Hernan G. Garcia

"Their exercises are unbloody battles, and their battles bloody exercises." - Flavius Josephus on Roman legionares

A Feeling For the Numbers

1 A Feeling for the Numbers in Biology: Your Turn

Over the semester, we will do many estimates about each biological phenomenon we address. To cement these skills, you will prepare two short estimates. Your first estimate will consist of a written vignette in the style of *Cell Biology by the Numbers*. You will present your second estimate in a 5-minute presentation at the end of the semester. Some examples of interesting estimates are

- How many proteins are in a viral capsid?
- What is the energy cost to a host cell in order to create a new virus after it has been infected?
- What is the cell-to-cell variability in the number of copies of the *lacZ* gene?
- What is the largest osmotic shock a cell can suffer without bursting?

Your first task is to write a short paragraph describing the estimate you're interested in writing a vignette about. Note that the objective at this point is not for you to have a finished estimate, but to have an outline of the calculation you plan to do so that we can give you feedback. Send this paragraph as an email to Hernan, Yang Joon and Jake by the homework due date.

2 The height of mountains on Mars

As a prelude to thinking about the buckling force of one-dimensional rods in the context of animal legs, we examined the physics behind mountain height. Using a simple relationship between mountain height and the weight of a column of rock and the stress needed to crush rock, we made an estimate of mountain height. In this problem, we use the observed height of the Olympus Mons on Mars which is 22 km high, to estimate the gravitational acceleration on the Red Planet.

(a) The estimate given in class was very hand wavy. In this part of the problem, let's do better. Specifically, the street fighter approach adopted in class argued that the mountain is a cylinder. Now consider a conical mountain with base radius R and height h and improve our earlier estimate. Explain the relationship you choose between h and R by commenting on some real world mountains.

(b) Given the scaling estimate for mountain height derived above, work out the ratio of mountain heights on Mars and those on Earth. Make sure to state all of your assumptions in constructing this ratio and then solve for g_{Mars}.

(c) As a second approach, use the observed g_{Mars} to make an estimate of the height of the tallest mountain on Mars.

A Feeling for Proteins

3 Protein Sequences: The Frances Arnold Estimate Problem

In a 2001 Bioengineering seminar at Caltech, Professor Frances Arnold made a startling remark that it is the aim of the present problem to examine. The basic point is to try and generate some intuition for the **HUGE**, **ASTRONOMICAL** number of ways of choosing amino acid sequences. To drive home the point, she noted that if we consider a protein with 300 amino acids, there will be a huge number of different possible sequences.

(a) How many different sequences are there for a 300 amino acid protein?

But that wasn't the provocative remark. The provocative remark was that if we took only one molecule of each of these different possible proteins, it would take a volume equal to five of our universes to contain all of these different *distinct* molecules.

(b) Estimate the size of a protein with 300 amino acids. Justify your result, but remember it is an estimate. Next, find an estimate of the size of the universe and figure out whether Frances was guilty of hyperbole or if her statement was on the money.

4 Post-Translational Modifications and "nature escape from genetic imprisonment"

In a very interesting article ("Post-translational modification: nature escape from genetic imprisonment and the basis for dynamic information encoding"), Prof. Jeremy Gunawardena discusses how we should think about post-translational modifications as a way of expanding the natural repertoire of the 20-letter amino acid alphabet. Similarly, Prof. Christopher Walsh (also at Harvard) wrote a whole book entitled "Posttranslational Modifications of Proteins: Expanding Nature's Inventory", again making the point that by adding chemical groups to proteins we can significantly change their properties.

(a) Provide at least one mechanistic idea about how adding a chemical group to a protein can alter its structure or function. Your answer should be offered in less than a paragraph, but should be concrete in its assertions about how these modifications change the protein. Why does Gunawardena refer to this process of post-translational modification as "escape from genetic imprisonment"?

(b) As a toy model of the combinatorial complexity offered by post-translational modifications, let's imagine that a protein has N residues that are able to be phosphorylated (NOTE: please comment on which residues these are - the answer is different for bacteria and eukaryotes). How many distinct states of the protein are there as a result of these different phosphorylated states? Make an approximate estimate of the mass associated with a phosphate group and what fraction of the total mass this group represents. Similarly, give some indication of the charge associated with a phosphate group. What ideas do you have about how we can go about measuring these different states of phosphorylation?

(c) In this part of the problem, we make a very crude estimate of the number of sites on a protein that are subject to phosphorylation. To do so, imagine that the protein is a sphere with N residues. How does the radius of that sphere depend upon the number of residues in the protein? Given that estimate, what is the number of residues that are on the surface? Given that number, what fraction of those are phosphorylatable? Remember, these are crude estimates. Work out these results for a concrete case of a typical protein with roughly 400 amino acids.

(d) Let's close out these estimates by thinking about a bacterial cell. If all 3×10^6 proteins in such a cell can be phosphorylated with the number of different phosphorylation states that you estimated above, how many distinct cells could we make with all of these different states of phosphorylation.

Bacterial Growth

5 Growth Curves and the Logistic Equation

In class, we discussed the exponential growth equation. This equation has been the basis for the study of microbiology for years (read, for example, F. Neidhardt, *Bacterial Growth: Constant Obsession with dN/dt*, J of Bacteriology 181:7405 (1999) provided on the course

website). If the number of cells is given by N and the growth rate is r, then this equation takes the form

$$\frac{dN}{dt} = rN. \tag{1}$$

We solved this equation in a variety of ways, both numerically and analytically, and found a solution given by

$$N(t) = N_0 e^{rt}, \tag{2}$$

where N_0 is the number of cells at $t = 0$.

(a) Of course, the solution shown above cannot be correct forever. For fast-growing *E. coli* estimate how long it would take for a single cell to produce enough progeny to cover the whole surface of the Earth.

A more realistic scenario is to account for the fact that, sooner or later, bacteria will run out of resources and halt their growth. For example, a liquid bacterial culture will saturate at a density of about 10^9 cells/ml. To account for these limited resources, we introduce a growth rate that depends on the number of cells, r_{new}

$$r_{new} = r\left(1 - \frac{N}{K}\right), \tag{3}$$

where K represents the maximum population size. Note that when N is very small compared to K, $r_{new} = r$ and growth is exponential. However, as N approaches K the growth rate will decrease. Thus, we get the so-called logistic equation

$$\frac{dN}{dt} = r_{new}N = rN\left(1 - \frac{N}{K}\right). \tag{4}$$

(b) What is the number of cells at which there is no growth and the population reaches steady state? Justify how you impose steady state on the logistic equation in order to figure out this number.

(c) In class, we wrote Python code to solve Equation 1 numerically. Modify your code to now solve the logistic equation. For reasonable choices of r and K, plot number of cells as a function of time for both exponential and logistic growth.

(d) Feel free to look at section "Computational Exploration: Growth Curves and the Logistic Equation" on page 103 of PBoC2.