MCB137L/237L: Physical Biology of the Cell
Spring 2020
Homework 2: A Feeling for *E. coli*
(Due 2/6/20)

Hernan G. Garcia

"Trying to understand molecular biology without numbers is like studying History without knowing Geography." - Prof. Marc Kirschner

A Feeling for the Numbers in Biology

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 **2/13 (when Homework 3 will be due)**.

2 Concentrations and absolute numbers in cells

(a) In class, we determined that a single molecule in *E. coli* will be at a concentration of about 1 nM. How many molecules are there in a a nucleus of the early embryo of the fruit fly, a yeast cell, and a fibroblast if that molecule is present at a concentration of 1 nM?
(b) Given that there are on the order of 3×10^6 proteins in *E. coli*, what is their average separation? How does this separation compare to the size of proteins themselves?
(c) How does the typical spacing of proteins inside the cell compare to the spacing of proteins you'd usually pipette into a test tube?

3 The Benjamin Franklin Problem

In his travels between America and Europe, Benjamin Franklin was subjected to the vicissitudes of the sea which led him to reflect on his reading of Pliny the Elder and claims of how oil was known to smooth the waves. Upon arriving in England, Franklin took the concept to the test. He tells us of his experience thus: "At length at Clapham where there is, on the common, a large pond, which I observed to be one day very rough with the wind, I fetched out a cruet of oil, and dropped a little of it on the water. I saw it spread itself with surprising swiftness upon the surface... the oil, though not more than a teaspoonful, produced an instant calm over a space several yards square, which spread amazingly and extended itself gradually until it reached the leeside, making all that quarter of the pond, perhaps half an acre, as smooth as a looking glass."

(a) Though Franklin himself never made the estimate (that was to await Lord Rayleigh), use Franklin's description of the experiment to work out the thickness of the oil film (the height of a lipid!) that covered the surface of Clapham common pond.

(b) Using a typical molecular mass for a lipid (say, 1000 g/mol), work out the number of lipid molecules that covered that surface of the pond and use that number to compute the area per lipid. How do your results compare to the modern values for the size of lipids?

The Central Dogma by the Numbers

4 Transcription by the Numbers

Solve problem 3.4 from PBoC2 (Figure 1). Note that the figure displays a so-called *Christmas Tree*. Here, RNA polymerase molecules are moving along the DNA and elongating mRNA. These RNA polymerase molecules show as dots along the DNA. As the mRNA is elongated it extends perpendicularly to the DNA like branches extending from a tree trunk. Ribosomes (also shown as dots) bind to these branches and begin translation.

5 Translation by the numbers

(a) Ribosomes can incorporate new aminoacids onto a nascent polypeptide chain at a rate of roughly 20 aa/s. Following logic similar to that we used for transcription, use the fact that

• **3.4 RNA polymerase and ribosomes**

(a) If RNA polymerase subunits β and β′ together constitute approximately 0.5% of the total mass of protein in an *E. coli* cell, how many RNA polymerase molecules are there per cell, assuming each β and β′ subunit within the cell is found in a complete RNA polymerase molecule? The subunits have a mass of 150 kDa each. (Adapted from Problem 4.1 of Schleif, 1993.)

(b) Rifampin is an antibiotic used to treat *Mycobacterium* infections such as tuberculosis. It inhibits the initiation of transcription, but not the elongation of RNA transcripts. The time evolution of an *E. coli* ribosomal RNA (rRNA) operon after addition of rifampin is shown in Figures 3.36(A)–(C). An operon is a collection of genes transcribed as a single unit. Use the figure to estimate the rate of transcript elongation. Use the beginning of the "Christmas-tree" morphology on the left of Figure 3.36(A) as the starting point for transcription.

(c) Using the calculated elongation rate, estimate the frequency of initiation off of the rRNA operon. These genes are among the most transcribed in *E. coli*.

(d) As we saw in the chapter, a typical *E. coli* cell with a division time of 3000 s contains roughly 20,000 ribosomes. Assuming there is no ribosome degradation, how many RNA polymerase molecules must be synthesizing rRNA at any instant? What percentage of the RNA polymerase molecules in *E. coli* are involved in transcribing rRNA genes?

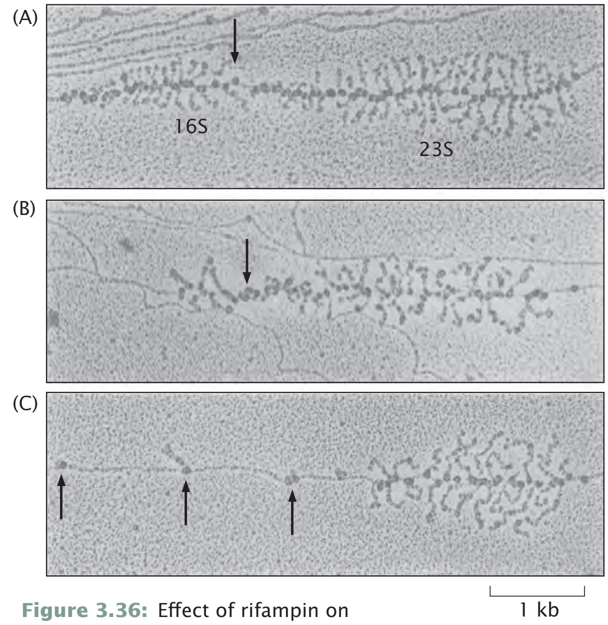

Figure 3.36: Effect of rifampin on transcription initiation. Electron micrographs of *E. coli* rRNA operons: (A) before adding rifampin, (B) 40 s after addition of rifampin, and (C) 70 s after exposure. No new transcripts have been initiated, but those already initiated are carrying on elongation. In (A) and (B) the arrow is used as a fiducial marker and signifies the site where RNaseIII cleaves the nascent RNA molecule producing 16S and 23S ribosomal subunits. RNA polymerase molecules that have not been affected by the antibiotic are marked by the arrows in (C). (Adapted from L. S. Gotta et al., *J. Bacteriol.* 20:6647, 1991.)

Figure 1: Problem 3.4 from PBoC2.

the footprint of the ribosome on the mRNA is roughly 35 nucleotides wide and determine the maximal rate at which new proteins could be synthesized.
(b) Given a lifetime of an mRNA molecule in *E. coli* of about 1 min, how many proteins can be produced off of this transcript during its lifetime? How does this compare to measurements of the so-called protein burst size in *E. coli* done by Cai *et al.* (article provided on the website)?

Bacterial growth revisited

6 Proteomic data on bacteria in different growth conditions

Read the paper by Schmidt and Heinemann and co-workers in which they use mass spectrometry to take the census of *E. coli* under a variety of different growth conditions. The outcome of this work was a census of the number of copies of roughly half of the proteins in this important bacterium.

(a) Using the data in the spreadsheet available with this homework, examine the numbers for the subunits of ATP-synthase. Write a short paragraph describing what ATP synthase is and what it does.

(b) In class, we posited that protein synthesis constitutes the largest source of ATP consumption, with 4 ATP equivalents required per amino acids incorporated. Make an estimate of the number of ATPs it takes to make a new cell. In light of the number of ATP synthases counted by Heinemann and his group, are there enough to make all the ATPs needed to build a cell?

(c) Comment on the units on the y-axis of Figure 2b of the Schmidt *et al.* paper. Specifically, justify those units in terms of what you know about the total number of proteins and the mass per protein. Do you think that the measurements pass the street fighters sanity check? Explain your conclusions.