BE/APh 161: Physical Biology of the Cell

Justin Bois

Caltech

Winter, 2019

This document was prepared at Caltech with financial support from the Donna and Benjamin M. Rosen Bioengineering Center.

2 Principles of estimation

As we discussed in the first lecture, one of our main aims in this course is to develop a sense of biological numeracy. A great way to do this is by performing estimates and back-of-the-envelope calculations about cells. Indeed, detail can often get in the way. For example, I may ask you if you prefer to fly or drive from Pasadena to San Francisco. You can hire a rideshare to Burbank, take a Southwest flight to Oakland, and then take the BART to San Francisco. The rideshare takes about 20 minutes, the flight about an hour, the BART takes about 20 minutes, and you have about an hour at the airport in Burbank and about 30 minutes at Oakland to get to the BART. In total, this is about 2.5 hours. Driving takes 5.5 hours without traffic, so you will choose to fly. Now, if we did this calculation trying to estimate how many minutes it takes for the Lyft driver to come, precisely how many minutes in baggage claim, etc., the calculation becomes cumbersome. Maybe more accurate, but you will still get more or less 2.5 hours to fly to San Francisco. It's easier, and more intuitive, to ignore all the little ins and outs. You end up with good intuition on how long things take nonetheless.

2.1 How many ribosomes in an *E. coli* cell?

Let's start learning about bacterial cells by performing a simple estimate: how many ribosomes are there in a single *E. coli* cell? We will take as given two pieces of data, a microscope image of an *E. coli* cell and an image of the growth rate, shown in Fig. 1, taked from *PBoC2* Fig. 3.8.

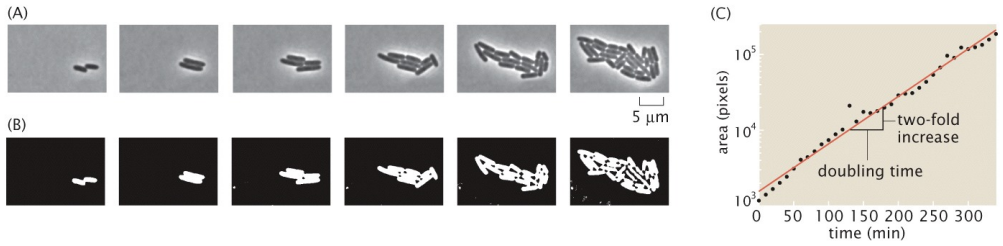

Figure 3.8 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Figure 1: A) Frames from a time lapse movie of bacterial growth with thresholding-based segmentation. B) A plot of the corresponding area of bacteria in the image as a function of time.

For copyright reasons, I do not show them here, but we will consider Fig. 3.8 of *PBoC2*. In particular, look at the image in Fig. 3.8B and the growth curve in Fig. 3.8C.

From the microscope image, we see that a single *E. coli* cell is about two microns long and one micron wide. From the growth curve, we see that the doubling time of *E. coli* in these conditions is about 45 minutes, or about 3000 seconds. Notice that I roughly estimated these values. 3000 seconds is 50 minutes, not 45, but we do not get bogged down in small differences like that.

At face this seems like a daunting task, estimating the number of ribosomes from only an image and a growth curve. But it is not so daunting if we divide the question up into smaller, more tractable (and less intimidating) questions. We can make a series of guesses to get us from what we know (the size of a cell and growth rate) to something else we *want to know* (the number of ribosomes).

Cell size (known).

1. **Estimate**: Cell density → Cell mass.

2. **Estimate**: Dry weight fraction $\rightarrow$ Cell dry mass.
3. **Estimate**: Fraction of dry mass that is protein $\rightarrow$ Cell protein mass.
4. **Known**: Cell division rate $\rightarrow$ Protein production rate by mass.
5. **Estimate**: Mass of amino acid $\rightarrow$ Protein production rate in units of AAs incorporated per second.
6. **Estimate**: Rate of ribosome function $\rightarrow$ Number of ribosomes.

Let's start estimating!

1. What is the cell density? A reasonable estimate is that it is close to that of water, which is 1 g/mL. We don't know the density, but we can **lie skillfully** to guess that it is the same as water. This is not a bad guess. To get the mass of a cell, we use our knowledge about its size. The volume of an *E. coli* cell is

$$V \approx \pi(1/2\ \mu\text{m})^2 \times 2\ \mu\text{m} \approx 1\ \mu\text{m}^3 = 1\ \text{fL}. \tag{2.1}$$

Note that the volume is closer to 1.5 μm^3, but for our rough estimates, we'll keep our numbers clear, and approximate this as one cubic micron, or one femtoliter. This is a nice number of keep around. **The volume of a single *E. coli* cell is about one femtoliter.**

Now that we have the volume and density, we compute the mass to be 10^{-12} grams, or one picogram, another useful number to keep in your head.

2. Now that we have the mass of the cell, how much of that is water? This is a bit tricky. You may have heard that your body is 80% water from popular lore. That's not actually that far off, and you could go ahead with your estimates taking that the *E. coli* cell is about 20% dry mass, or 0.2 pg.

Another option I considered it to recall my training from chemistry classes where I looked at closest packing of spheres. I remember that the fractional void volume of closest packed spheres is $1 - \pi/\sqrt{18}$, or about 25%. (I have no idea why that stuck in my head.) That's for closest packed spheres, but things need to diffuse around in the cell, so the void volume is probably more than that, say two or three times that. So, we'll say that the solid material takes about one-third of the total cell volume. 33% is not too different from 20%, and we'll use 33% going forward.

This tactic in our educated guesswork, where we use any knowledge that happens to be in our head, is called **guerrilla warfare**.

3. How much of this dry mass is actually protein? In a cell, proteins *do* most things and have longer lifetimes than, say, RNAs. So, let's guess that about half of the dry mass is protein, so about 1/6 of the total cellular mass is protein, or about 15%. This gives a protein mass of 0.15 pg.

You might think that this is nuts. I'm making guesses seemingly off the top of my head. But this exposes an important principle of educated guess work: **Fear not!**, lest you be paralyzed. Sometimes you need to make these guesses to move forward. Just do it, and **cross-check** later.

4. From the plot of the growth curve, we estimated that *E. coli* under those conditions divided every 45 minutes. Thus, each cell needs to produce 0.15 pg of protein every 45 minutes. Forty-five minutes is about 3000 seconds, so the bacterium needs to produce 5×10^{-5} pg of protein every second.

For convenience, let's work in mass units of Daltons. We have

$$5 \times 10^{-5} \text{ pg/s} \times \frac{1 \text{ g}}{10^{12} \text{ pg}} \times \frac{6 \times 10^{23} \text{ Da}}{1 \text{ g}} \approx 3 \times 10^{7} \text{ Da/s}. \tag{2.2}$$

5. How many amino acids are there in 3×10^7 Daltons worth of protein? A typical amino acid has a nitrogen, two carbons, two oxygens, and a side chain. The nitrogen contributes 14 Daltons, the carbons 24, and the oxygens 32, for a total of 70 Daltons, exclusive of hydrogens and side chains. We'll estimate that the side chains, on average, bring the mass of each amino acid up to about 100 Da. So, the bacterium incorporates about 3×10^5 amino acids into proteins per second.

6. We are now left to estimate the rate at which ribosomes function. This is a hard one to guess. If we guess it to be too slow, the cell will fill up with ribosomes, which sets a hard lower bound on this guess. To make this guess, I will use the fact that diffusion-limited chemical reactions between proteins tend to proceed with rate constants around 100 to 1000 s^{-1}. There are several reactions that have to happen to add an amino acid to a protein, including diffusion of the tRNA into the pocket of the ribosome, formation of covalent bond, vacation of tRNA, moving the DNA strand forward etc. These are also big complexes, so I will estimate the rate to be an order of magnitude slower, about 10 AAs per second.

So, if the bacterium incorporates 3×10^5 amino acids into protein per second, and each ribosome does this at a rate of about 10 amino acids per second, there are about 30,000 ribosomes in an *E. coli* cell. And we have arrived at the estimate we sought.

2.2 Principles of estimation

We have performed an estimate of the number of ribosomes in an *E. coli* cell. In doing so, we have navigated some seemingly dangerous waters, but in the end emerged with an estimate quite close to the reported value.

As you do more estimates, there are some principles of estimation to keep in mind.

1. **Fear not!**
   It is all too easy to be paralyzed and be afraid to make an estimate because it is too crude or you are not sure enough. *Just do it!* You can always come back later and refine.
2. **Divide and conquer.**
   We just saw that it is easier to break the problem down into smaller, easier estimates. This is a key strategy for tackling what might be at first glance a really tough quantity to estimate.
3. **Talk to your gut.**
   When you are making estimates, ask yourself, "Does this feel right?". Somehow your collection of experiences in your life can help you have a good gut feel for things.
4. **Lie skillfully.**
   Do you know the density of *E. coli*? Probably not. But you can lie and say it's the same as water. And this is a good lie, because it's not far from the truth, and you know it's not far from the trust. This is a good, skillful lie. Such lying will help you with your estimates.
5. **Guerrilla warfare.**
   Use everything available to you!

6. **Cross-check.**
   After making an estimate, try making it using another divide-and-conquer strategy. If the two estimates do not match, it is time to check what may have messed things up. Such inconsistencies are a great way to find flawed (and good) logic.

2.3 More practice

It's common to have questions pop in your mind when making estimates. Now that you have in your mind the number of ribosomes in an *E. coli* cell, try approaching these questions.

1. What fraction of the protein material in an *E. coli* cell is made out of ribosomes?
2. How many mRNA transcripts are there in an *E. coli* cell at any given time? With this number, how many mRNA molecules are there per gene?
3. How does the mass of mRNA in an *E. coli* cell compare to that of DNA?
4. How does the mass of ribosomal RNA compare to that of mRNA and DNA?

There are many ways to approach these practice problems. Here are the approaches I took.

What fraction of the protein material in an *E. coli* cell is made out of ribosomes? A typical protein is about 300 amino acids, giving a mass of about 3×10^4 Da. Ribosomes contain about 50 proteins, so their mass is about 10^6 Da. (The real value is about three times this.) With 30,000 ribosomes, this amounts to 3×10^{10} Da in ribosome protein mass. We already worked out that the total protein mass is about 0.15 pg, which is about 10^{11} Da. By our estimate, then, a third of protein in an *E. coli* cell are ribosomes. If we take the actually molecular mass of ribosomes, we get that almost all protein is ribosomes. So, the cell is basically just a ribosome factory during optimal growth conditions!

How many mRNA transcripts are there in an *E. coli* cell at any given time? We can approach this problem from above and from below. As an upper bound, we can imaging that each mRNA molecules is a single ribosome attached to it. This would mean we have 30,000 mRNA molecules as an upper bound. As a lower bound, we can imagine that each mRNA molecule is completely covered in ribosomes. To perform the calculation, then, we need to know the width of a ribosome and the length of an mRNA transcript.

A ribosome has a "volume" of about 50 proteins, so it has a diameter of about $\sqrt[3]{50} \approx 4$ proteins. Proteins are typically a few nanometers across, so a ribosome is about 12 nanometers across. At least half of the ribosome is rRNA, which we have not yet considered so we'll double this number to about 20 nm.

To compute the length of an mRNA transcript, we note that the stack height of RNA

is about 0.4 nm. If a typical protein has 300 amino acids, this means there are about 900 mRNA bases, for a total length of about 500 nm.

So, if the ribosomes completely cover the mRNA, we have a total length of $30,000 \times 20$ nm = 600,000 nm worth of mRNA. This amounts to about 1000 mRNA molecules. So, there is somewhere between 10^3 and 10^4 mRNA molecules. We can take the geometric mean to get that we have about 3,000 mRNA molecules in an *E. coli* cell. BNID 100064 says that there are about 1400, closer to the lower bound, suggesting that the mRNA molecules are densely decorated with ribosomes, which is what we would expect for a rapidly (efficiently) growing cell.

Since *E. coli* has about 5000 genes, there are only about 0.2 copies of mRNA per gene in *E. coli*.

How does the mass of mRNA in an *E. coli* cell compare to that of DNA?
RNA and DNA have similar molecular masses per base, with mRNA being a bit heavier. We'll take them to have the same per base mass. The genome is about 4.6 million base pairs, or about 10 million DNA bases. If we have 3,000 mRNA molecules, each with about 900 bases, we have about 3 million RNA bases. So, the mass of DNA is about three times than of mRNA. Here we have neglected multiple copies of the genome due to multiple replication forks.

How does the mass of ribosomal RNA compare to that of mRNA and DNA?
If we compare total RNA mass to DNA mass, we need to consider also the rRNA. Each ribosome is about half rRNA and has a molecular mass of about 3×10^6 Da. So, there is about 10^6 Da of rRNA in a ribosome, giving about 3×10^{10} Da of rRNA in the 30,000 ribosomes in the cell. Each base is about 300 Da, giving a total of about 10^8 bases worth of rRNA. This is a couple orders of magnitude bigger than the amount of mRNA in the cell, and an order of magnitude bigger than the DNA mass.