MCB137L/237L: Physical Biology of the Cell
Spring 2025
Homework 1: Biological Numeracy
(Due 1/28/24 at 2:00pm)

Hernan G. Garcia

Last updated January 21, 2025

"The greats weren't great because at birth they could paint. The greats were great cause they paint a lot." - Macklemore

Homework assignments in MCB137L/237L

Whether it is in the context of professional sports, art, or science, it pays to practice. The idea of the homework assignments during our course is to give you a venue to get your 10,000 hours of calculations and estimates in as a means to become proficient in the mathematical and physical modeling of living systems. Sometimes, the problems will ask you to redo a derivation we did in class in a new way, and sometimes they will propose a whole new biological phenomenon to attack using some of the tools we explored in class. Further, we will use some of the homework problems to introduce and elaborate on new concepts. Regardless, if you spend more than five hours on a homework set, it means that you should come to office hours. Make sure to start working on your problems early on!

The objective of this homework set is to get a feeling for the numbers in whatever problem you're considering in biology. Just like you always need to check the units in your calculations, a more subtle sanity check of your theoretical results stems from having some expectation about the order of magnitude you will obtain.

This first problem set involves a number of challenges in order-of-magnitude thinking. When doing street fighting estimates, the goal is to do simple arithmetic of the kind that all numbers take the values 1, few (f) or 10. few $\times$ few = 10, etc. Please do not provide estimates with multiple "significant" digits that are meaningless. Figure 1 presents a good set of rules to figure out how many significant digits to use when reporting on your estimates. Be thoughtful about what you know and what you don't know. You may use the Bionumbers website (`http://bionumbers.hms.harvard.edu/`) to find key numbers (examples are masses of amino acids (BNID 104877) and nucleotides (BNID 103828), the speed of the ribosome (BNID 100059), etc.), but please provide a citation to the Bionumber of interest as shown above. However, for

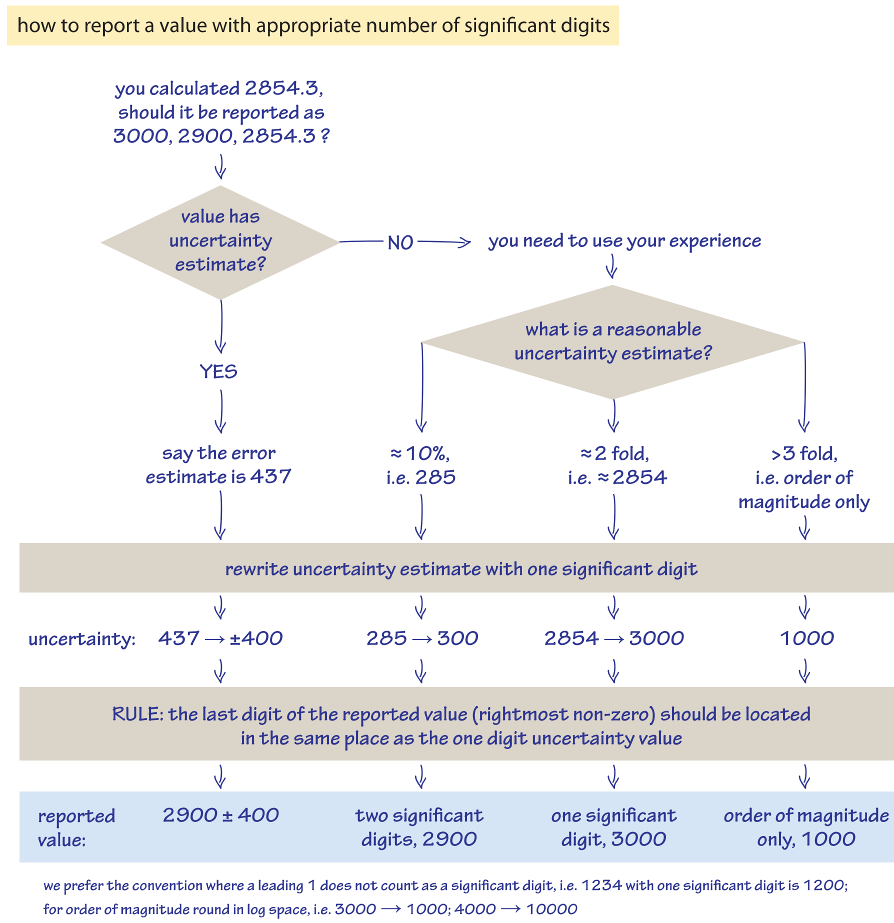

Figure 1: A flow chart to help determine how to report values with an appropriate number of significant digits. Adapted from Milo, R and Phillips, R (2016) Cell Biology by the Numbers, Garland Science

many of these problems the essence of things is to do simple estimates, not to look quantities up. In particular, if in doubt, use the square root rule

$$x_{guess} = \sqrt{x_{low}\, x_{high}}, \tag{1}$$

which instructs us to take a lower and upper bound guess and then to take their geometric mean (which is the same as averaging their exponents).

Sometimes, the problems will be drawn directly from the 2nd edition of Physical Biology of the Cell (PBoC or PBoC2). In that case, I'll make the effort to scan the problems and include them as a figure. However, some of those problems might refer to information inside the book, which I will not scan. As a result, you might want to just get the book.

Homework submission

Gradescope will be used to submit and grade your homework. We will create two submissions for weekly homework (one for written pdf, the other one for submitting a zip file for your code). When you have to write Python code in order to make plots, you don't need to include your actual code in the pdf document you submit. However, you need to submit all the code you used to generate the plots in a separate zip file. All submitted code should be well documented. Besides this being useful every time you code, documenting will make it possible for us to follow your reasoning. All plots you generate need to have axes and lines that are clearly labeled. Please submit both pdf and original code before the deadline.

Finally, remember to write each problem on a different piece of paper so that you can upload them independently to Gradescope. This will make it easier for us to grade them.

How to join Gradescope

1. Go to website: https://www.gradescope.com
2. Create an account
3. Add class with entry code: **6J232D**
4. Please update your "Student ID" in the account settings

1 I wonder

Give three thoughtful sentences that start with the two words "I wonder." Make sure that these "I wonder" sentences concern the nature of the living world writ large.

2 Benjamin Franklin and Molecular Dimensions

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 in an experiment like that shown in Figure 2), use Franklin's description of the experiment to

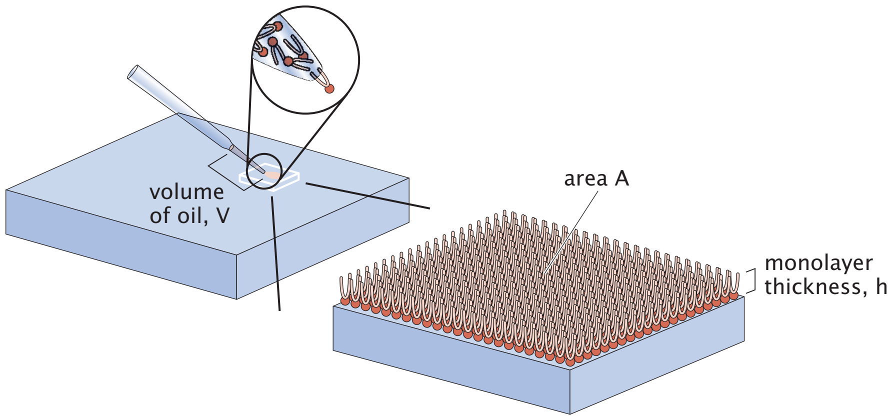

Figure 2: Putting oil on water to measure molecular dimensions. Here we see that the lipid molecules form a monolayer.

TABLE I.

Preliminary Measurements of Cross-Sections and Lengths of Molecules.

Substance	Formula	I. Cross-section. Sq. cm.	II. $\sqrt{\text{Cross. sec.}}$ Cm.	III. Length. Cm.	IV. Length per carbon atom.
Palmitic acid	$C_{15}H_{31}COOH$	21×10^{-16}	4.6×10^{-8}	24.0×10^{-8}	1.5×10^{-8}
Stearic acid	$C_{17}H_{35}COOH$	22×10^{-16}	4.7×10^{-8}	25.0×10^{-8}	1.39×10^{-8}
Cerotic acid	$C_{25}H_{51}COOH$	25×10^{-16}	5.0×10^{-8}	31.0×10^{-8}	1.20×10^{-8}
Tristearin	$(C_{18}H_{35}O_2)_3C_3H_5$	66×10^{-16}	8.1×10^{-8}	25.0×10^{-8}	1.32×10^{-8}
Oleic acid	$C_{17}H_{33}COOH$	46×10^{-16}	6.8×10^{-8}	11.2×10^{-8}	0.62×10^{-8}
Triolein	$(C_{18}H_{33}O_2)_3C_3H_5$	126×10^{-16}	11.2×10^{-8}	13.0×10^{-8}	0.69×10^{-8}
Trielaidin	$(C_{18}H_{33}O_2)_3C_3H_5$	120×10^{-16}	11.0×10^{-8}	13.6×10^{-8}	0.72×10^{-8}
Cetyl palmitate	$C_{15}H_{31}COOC_{16}H_{33}$	23×10^{-16}	4.8×10^{-8}	41.0×10^{-8}	2.56×10^{-8}
Myricyl alcohol	$C_{30}H_{61}OH$	27×10^{-16}	5.2×10^{-8}	41.0×10^{-8}	1.37×10^{-8}

Figure 3: Values for the size of lipids obtained by Irving Langmuir in 1916 using the so-called Langmuir trough, earlier used to great advantage by Agnes Pockels.

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 - give an order of magnitude justification of this rule of thumb), 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?

3 Composition of a cell

Here we are going to do a rough atomic census of living material by thinking about the principal ingredients of a cell. To get a sense of the chemical makeup of the dry mass of a cell, we are going to focus only on proteins and nucleic acids.

(a) Provide a simple and clean estimate for the volume and mass of a typical bacterium such as *E. coli*.

(b) One of the key rules of thumb we will invoke over and over again is a knowledge of the concentration of one molecule per *E. coli* cell. Using the volume from part (a), work out a simple estimate for the concentration of 1 molecule per *E. coli* cell. Remember that we are in street-fighting mode and thus your answer should be 1, few or 10 in nM, μM, mM or M.

(c) Assume that 1/3 of the mass of a bacterium is dry mass and for simplicity, we ascribe all of that dry mass either to proteins or nucleic acids. We will take our elemental composition of a "typical" amino acid to be $N_1C_5O_2H_8$ and a "typical" nucleotide to be $P_1N_5O_7C_{10}H_{14}$. Given that roughly half the dry mass of the cell is protein, work out the number of proteins and hence, the number of amino acids per cell.

(d) As an alternative approach to estimating the total number of proteins in *E. coli*, assume that the bacterium is tightly packed with proteins (think of golf balls in a bathtub). How does this compare to the estimate from part (c)?

(e) Work out the number of nucleotides in the genome of our bacterium of interest.

(f) Finally, figure out how many ribosomes are needed, translating at roughly 15 aa per second to translate all of those proteins required to make a new cell. How many nucleotides are present in the ribosomal RNA making up all of these ribosomes?

(g) Given all of these numbers from the rest of this problem, you are now able to work out the overall composition of a cell. Provide an approximate formula for the stoichiometry of a bacterium.

4 To build a cell

Minimal growth medium for bacteria such as *E. coli* includes various salts with characteristic concentrations of mM and a carbon source. This carbon source is typically glucose and it is used at 0.2% (a concentration of 0.2 g/100 mL).

(a) Make an estimate of the number of carbon atoms it takes to make up the macromolecular contents of a bacterium such as *E. coli*.

(b) Make an estimate of the number of nitrogen atoms it takes to make up the macromolecular contents of a bacterium such as *E. coli*.

(c) How many cells can be grown in a 5 mL culture using minimal medium before the medium exhausts the carbon? Note that this estimate will be flawed because it neglects the *energy* cost of synthesizing the macromolecules of the cell. Similarly, given that the recipe for minimal media requires ammonium chloride NH_4Cl at a concentration of 100 mM, how many cells can be grown in a 5 mL culture using minimal medium before the medium exhausts the nitrogen?

(d) In rapidly dividing bacteria, the cell can divide in times as short as 1200 s. Make a careful estimate of the number of sugars (glucose) needed to provide the carbon for constructing the macromolecules of the cell during one cell cycle of a bacterium. Use this result to work out the number of carbon atoms that need to be taken into the cell each second to sustain this growth rate.

(e) These problems are intended to get you thinking about the wondrous process whereby cells convert a clear liquid with simple chemical ingredients into biomass as shown in Figure 4. Amazing! Now, work out an estimate related to the volume of the headspace you see in Figure 4 which has oxygen available for cell growth. Specifically, if 6 O_2 molecules are consumed for every sugar, make a simple estimate of the required volume of headspace needed to sustain cell growth. Note that our estimate about O_2 usage is crude and sloppy. To really do this carefully, we need to acknowledge the use of glucose both in providing building materials (i.e. carbon skeletons) as well as the energy needed to synthesize a cell. The estimate we do here is intended to give an impression of the magnitudes, and specifically to get a sense of the aeration requirements when we do a liquid culture growth procedure.

Figure 4: Growth of *E. coli* in rich media. The tube on the left shows roughly 5 mL of growth media just after inoculation. The tube on the right shows such media after saturation due to exponential cell growth and division.

5 Sizing up the Central Valley

California's Central Valley is one of the most potent agricultural regions in the world. In this problem, you are going to evaluate many of the key factors associated with its enormous productivity without any data aside from a single satellite image of the region as shown in Figure 5. Note that the key point here (and what you will be graded for if you care about such things) is the logical flow of your estimates, not the particular numerical values you found.

(a) Water usage. Using what you know about watering and the growth of plants, make an estimate of the amount of water used to irrigate the agriculture of the Central Valley.

(b) Nitrogen usage. Since the beginning of the twentieth century, we have doubled the number of occupants that can be fed on earth as a result of the Haber-Bosch process and the synthetic fixation of nitrogen. In this part of the problem, begin by estimating the number of kilograms of biomass per square meter that is produced per year. From that number, figure out how many kilograms of nitrogen are contained per square meter of biomass. Then, make an estimate of how much fertilizer is used for each square meter and hence for the entirety of the Central Valley.

(c) Pesticide usage. Undertake an estimate similar to that in the first two parts of the problem to figure out how much pesticide is used on the Central Valley every year.

(d) Do NOT do this part until you have done parts (A) - (C). Look up some source of data on each of these three questions and compare your results to the data. Please do not redo your estimate.

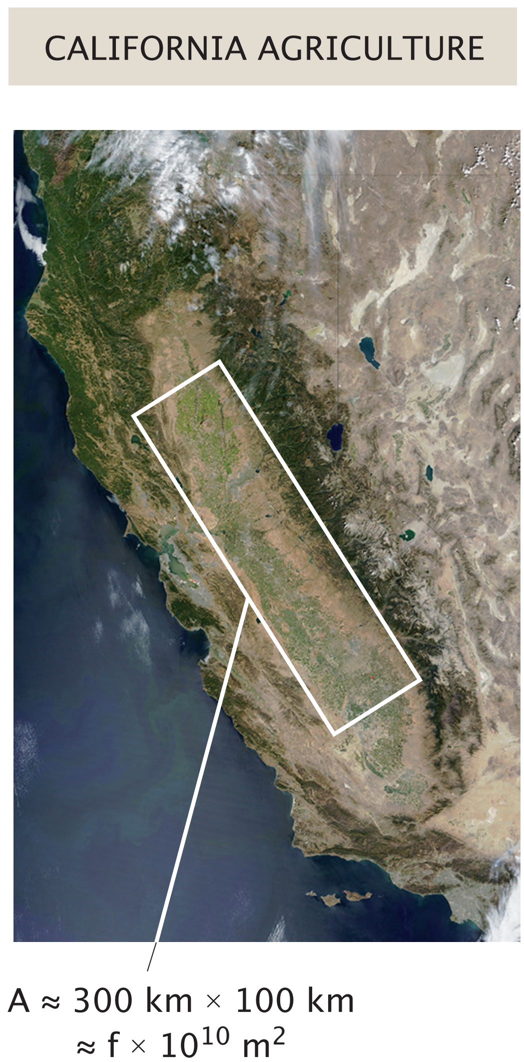

Figure 5: Satellite image of California's Central Valley.