BE/APh161: Physical Biology of the Cell
Homework 1
Due Date: Wednesday, January 13, 2021

"A physics that has no place for life is as impoverished as would be a biology not informed by chemistry. The study of life as a natural phenomenon, a fundamental feature of the universe, must not be allowed to slip into the black hole of departmental tribalism." - Franklin Harold, *The Way of the Cell*

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. 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 many of these problems the essence of things is to do simple estimates, not to look quantities up.

Extra Credit. Provide comments on chap. 2, "Setting the Scales of Living Things" of the upcoming third edition of *Physical Biology of the Cell*. Note that this is an unfinished draft of the chapter. Figure placements are not necessarily correct and there are still a number of internal discussions amongst the author team about how to finish things off. We are especially interested in mistakes, flaws in logic, confusing figures, unclear discussions, etc., but are happy to entertain comments at all scales. This extra credit will constitute an additional 15% on your score on the homework.

1. 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

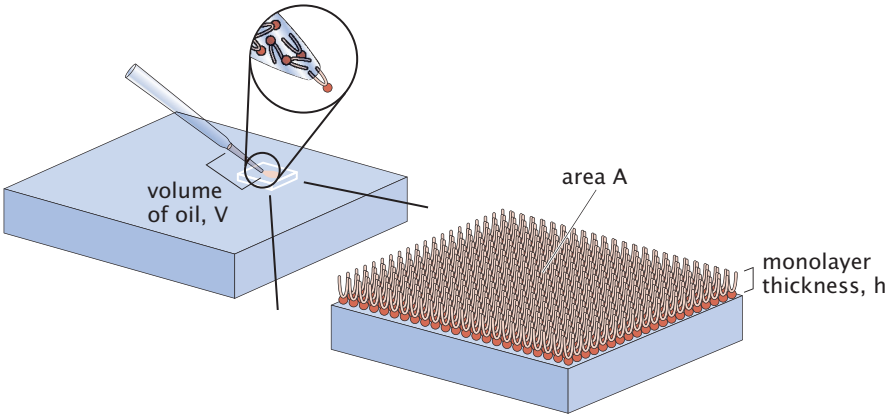

Figure 1: Putting oil on water to measure molecular dimensions. Here we see that the lipid molecules form a monolayer.

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?

2. Street fighting your way to the ribosome density.

One of the most important molecular assemblies in the cell is the ribosome. The number of ribosomes per cell dictates how fast cells can grow. *E. coli* growing with a division time of 24 minutes have 72,000 ribosomes per cell, and slow growing *E. coli* with a division time of 100 minutes have a factor of

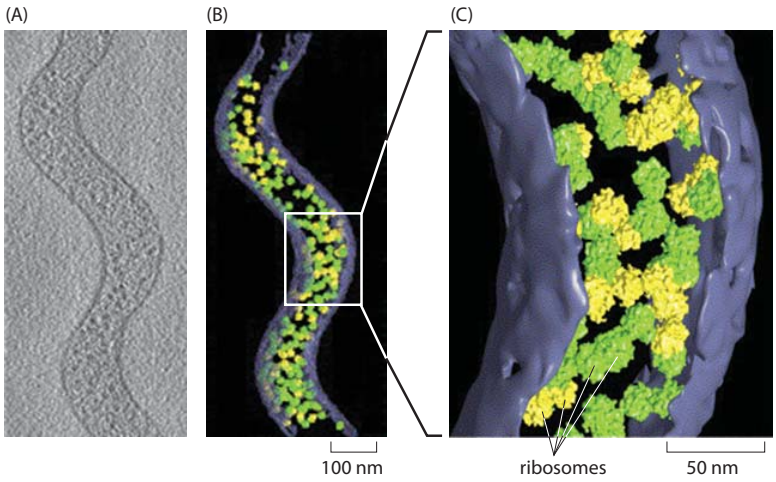

Figure 2: Cryo EM study of a bacterial cell. These images are of the tiny bacterium, *Spiroplasma melliferum*. Using algorithms for pattern recognition and classification, components of the cell such as ribosomes were localized and counted. (A) Single cryo-electron microscopy image. (B) 3D reconstruction showing the ribosomes that were identified. Ribosomes labeled in green were identified with high fidelity while those labeled in yellow were identified with intermediate fidelity. (C) Close up view that you should use to make your count. Adapted from JO Ortiz *et al.*, J. Struct. Biol. 156, 334-341 (2006).

ten fewer ribosomes with a count of ≈ 6800 ribosomes. In this problem, we will use our street fighting skills to explore the ribosomal density in another organism as shown in Figure 2, and then see how well our results from the electron microscopy study square with the numbers quoted above. By examining the figure, make an estimate of the number of ribosomes per μm^3 and compare that result to the numbers quoted for *E. coli* above.

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. Assume that 1/3 of the ≈ 1 pg 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. Then, work out the number of nucleotides in the genome of our bacterium of interest. Finally, figure out how many ribosomes are needed, translating at roughly 15 aa per second to translate all of those proteins. How many nucleotides are present in all of these ribosomes? Given all of these numbers, 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 $0.2\,\mathrm{g}/100\,\mathrm{mL}$).

(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) 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.

(c) 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.

(d) 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 3. Amazing! Now, work out an estimate related to the volume of the headspace you see in Figure 3 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 3: 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 4. 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.

6. RNA Polymerase and Rate of Transcription.

One of the ways in which we are trying to cultivate a "feeling for the organism" is by exploring the processes of the central dogma. Specifically, I want you to have a sense of the number of copies of the key molecular players in the central dogma as well as the rates at which they operate. Further, I argue that it is critical you have a sense of *how* we know these numbers.

(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.)

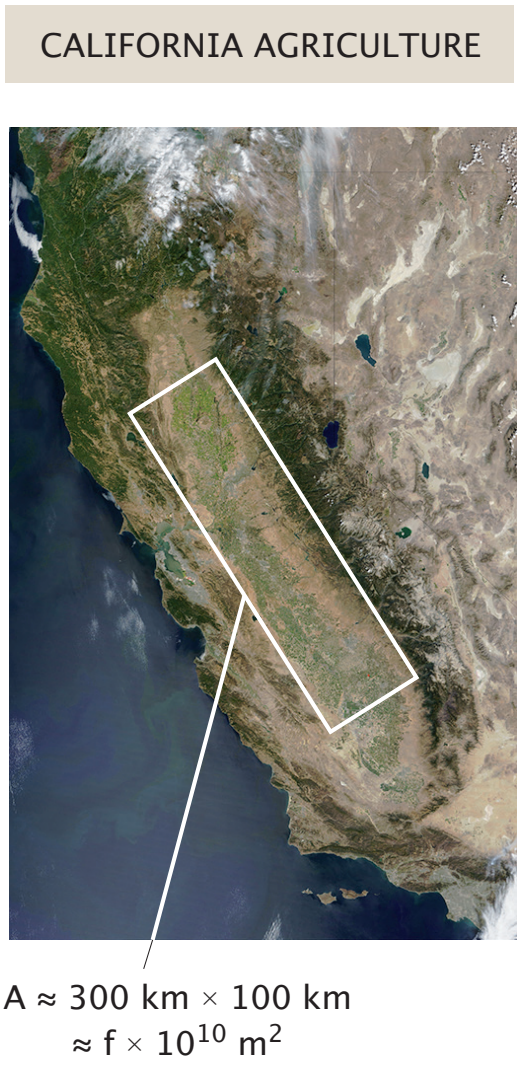

Figure 4: Satellite image of California's Central Valley.

(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 Figure 3.29(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 3.29(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 amongst the most transcribed in *E. coli*.

7. A feeling for the complete blood count (CBC) test.

Typical results for a complete blood count (CBC) are shown in Table 1. Assume that an adult has roughly 5 L of blood in his or her body. Based on these values estimate:

(a) the number of red blood cells.

(b) the percentage in volume they represent in blood.

(c) their mean spacing.

(d) the total amount of hemoglobin in the blood.

(e) the number of hemoglobin molecules per cell.

(f) the number of white blood cells in the blood.

8. The pandemic elephant in the room.

We are living through a global pandemic that has changed all of our lives in far reaching ways. As a result, each week, we will have at least one problem that reminds us of the pandemic, and asks us to think about it quantitatively. in this problem, we are going to explore the mass of an individual SARS-CoV-2 virion, the total mass of such viruses within a given individual at the

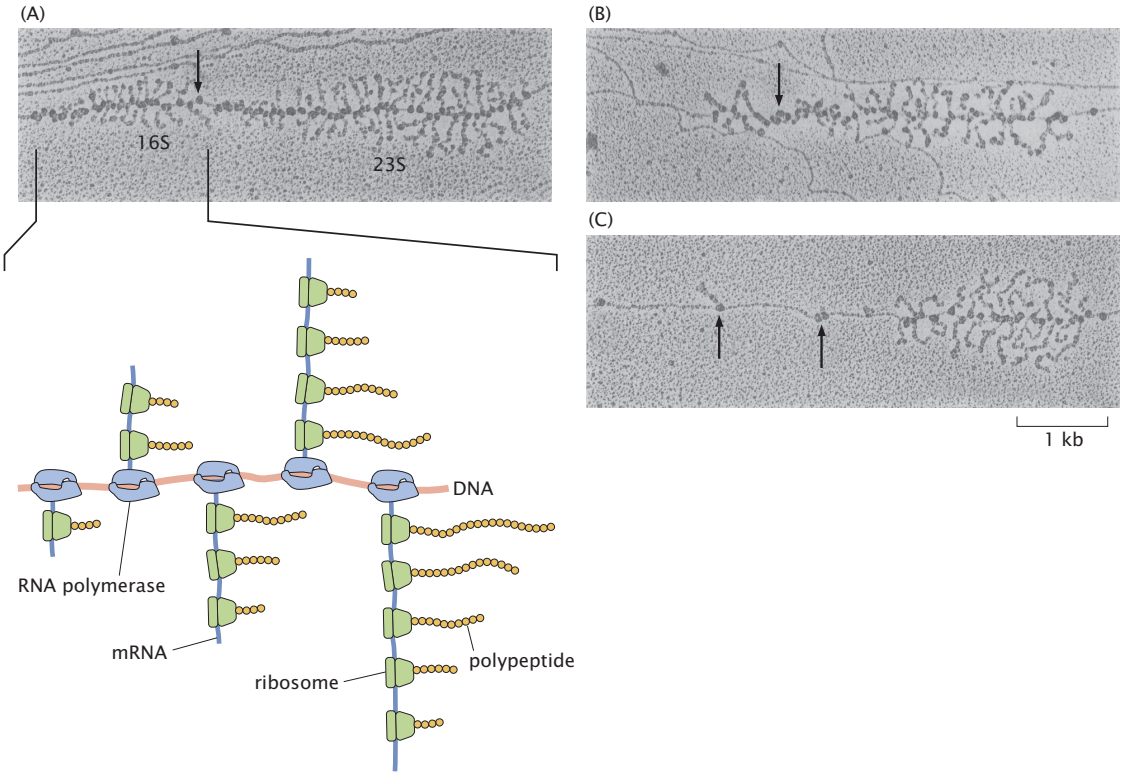

Figure 5: Effect of rifampin on transcription initiation. Electron micrographs of *E. coli* rRNA operons: (A) before adding rifampin, (B) 40 s after addition of rifampin, and (C) 70 s after exposure. No new transcripts have been initiated, but those already initiated are carrying on elongation. In parts (A) and (B) the arrow 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 part (C). (Adapted from L. S. Gotta et al., *J. Bacteriol.* 20:6647, 1991.)

Test	**Value**
Red blood cell count (RBC)	Men: $\approx(4.3\text{–}5.7) \times 10^6$ cells/μL
	Women: $\approx(3.8\text{–}5.1) \times 10^6$ cells/μL
Hematocrit (HCT)	Men: $\approx(39\text{–}49)\%$
	Women: $\approx(35\text{–}45)\%$
Hemoglobin (HGB)	Men: $\approx(13.5\text{–}17.5)$ g/dL
	Women: $\approx(12.0\text{–}16.0)$ g/dL
Mean corpuscular hemoglobin (MCH)	$\approx(26\text{–}34)$ pg/cell
MCH concentration (MCHC)	$\approx(31\text{–}37)\%$
Mean corpuscular volume (MCV)	$\approx(80\text{–}100)$ fL
White blood cell count (WBC)	$\approx(4.5\text{–}11) \times 10^3$ cells/μL
Differential (% of WBC):	
Neutrophils	$\approx(57\text{–}67)$
Lymphocytes	$\approx(23\text{–}33)$
Monocytes	$\approx(3\text{–}7)$
Eosinophils	$\approx(1\text{–}3)$
Basophils	$\approx(0\text{–}1)$
Platelets	$\approx(150\text{–}450) \times 10^3$ cell/μL

Table 1: Typical values from a CBC. (Adapted from R. W. Maxwell, Maxwell Quick Medical Reference, Tulsa, Maxwell Publishing Company, 2002.)

peak of their infection and the total mass of all the SARS-CoV-2 viruses on the planet.

(a) Given the roughly ≈ 100 nm diameter of a single SARS-CoV-2 virion, work out a simple estimate for its mass. What fraction of that mass corresponds to the genome? To answer the latter question, use simple rules of thumb for the mass of a nucleotide and use the fact that this virus is a single-stranded RNA virus with a roughly ≈ 30 kb genome.

(b) There are a number of cell types in different tissues that are susceptible to infection by SARS-CoV-2. For our purposes, we are going to focus on the most massive such tissue, namely, the lungs. There are several different assays for measuring the viral load within an infected individual. One method is to use RT-PCR to amplify their nucleic acid content with the result that there are between $10^6 - 10^8$ RNA copies per gram of lung tissue. Alternatively, infectious virions are measured by using cells in tissue culture and figuring out at what concentration of viruses, half of the tissue culture cells will be infected, the so-called TCID50 (tissue-culture infectious dose). Samples from lung tissue yield the range of $10^2 - 10^4$ TCID50 per gram of lung material. Using these results, estimate the total number of virions in the lung and comment on the difference between the RNA-based assay and the infection assay. Given these numbers, what is the total mass of viruses within an infected individual at the peak of their infection? To the extent that our estimate is correct, what fraction of virions are actually infectious?

(c) Use the results of the previous two parts of the problem to estimate the total mass of all the SARS-CoV-2 viruses that have been present in the human population since the beginning of the pandemic.