BE/APh161: Physical Biology of the Cell
Homework 1: Solutions
Due Date: Wednesday, January 15, 2020

"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*

When doing street fighting estimates, the goal is to do simple arithmetic of the kind that all numbers are 1, few 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.

1. 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 1. 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.

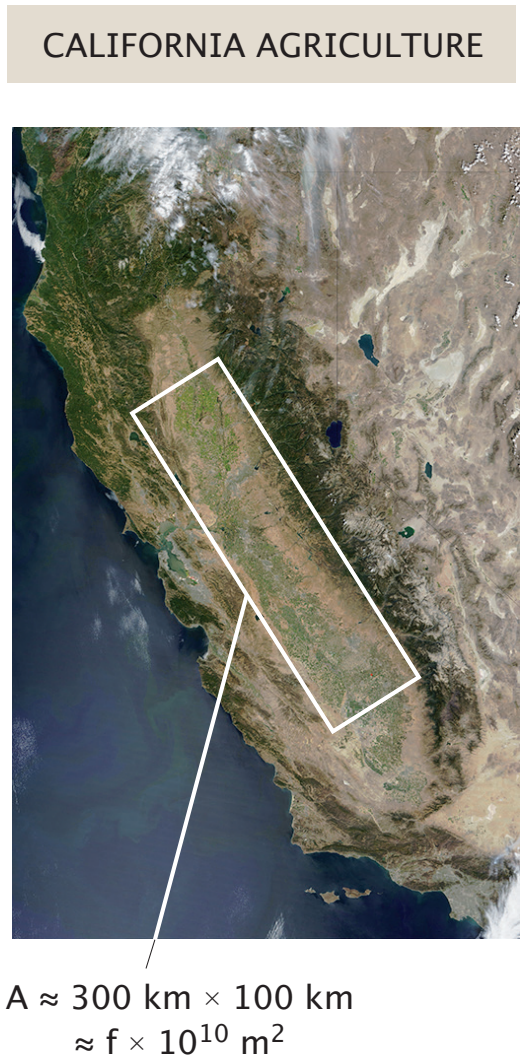

Figure 1: Satellite image of California's Central Valley.

(a) We will assume that winter (December-February) is too cold for the crops to grow. Due to the weather of California, there is limited rainfall in the area, so we will assume that all the water that crops use to grow comes from irrigation. From the satellite image, we know that the size of Central Valley is around 10^{10} m^2. We will assume that all the regions are used for agriculture to simplify our calculation. Next, we would like to estimate how many liters of water is needed everyday to irrigate the crop. We would estimate this number based on our daily experience taking care of flowers at home. From our experience, a typical size of a flowerpot is 20 cm $\times$ 20 cm and we would need a cup of water (250 mL) everyday to

irrigate the flower. That would give us:

$$\begin{aligned}\text{Water needed to irrigate crops in a unit area per day} &= \frac{0.25 \text{ L/day}}{20 \text{ cm} \times 20 \text{ cm}} \\ &= \frac{0.25 \text{ L/day}}{0.04 \text{ m}^2} \\ &\approx 5 \text{ L/m}^2 \cdot \text{day}. \end{aligned} \tag{1}$$

Then, we can estimate the total amount of water needed every year as:

$$\underbrace{10^{10} \text{ m}^2}_{\text{Central Valley area}} \times 5 \,\text{L/m}^2 \cdot \text{day} \times 30 \text{ days/month} \times \underbrace{9 \text{ months/year}}_{\text{crop season}} \approx 10^{13} \,\text{L/year}. \tag{2}$$

Just to help you get a sense of how much water this is, the average volume of water in Lake Tahoe is 37 trillion gallons, which is roughly 1.4×10^{14} L (https://www.fs.usda.gov/main/ltbmu/about-forest/about-area). So, the amount of water we estimated above is about 1/10 of the volume of Lake Tahoe.

(b) The biomass produced depends on the type of plant being grown, so we will only estimate the order of magnitude. We can estimate biomass per square meter based on everyday experience and specific examples. Take watermelon as an example; we can harvest a few watermelons per m^2 and each watermelon weighs about a few kg, so we can use the trick of few $\times$ few $\approx$ 10 to get:

$$\text{Biomass per square meter} \approx 10 \text{ kg/m}^2. \tag{3}$$

To estimate the amount of nitrogen contained in plants, we need to better understand the plant composition. Plants are composed of water, carbon-containing organic, and non-carbon-containing inorganic substances. We know that approximately 95% of plant is made of water, so less than 5% of biomass is composed of organic and inorganic substances.

Nitrogen is a critical component of amino acids in protein. To estimate the amount of nitrogen contained in the remaining biomass (5%), we will assume that it is composed of amino acids. Considering the atomic composition of amino acids, we can say that on average they contain 2 oxygen (16 g/mol), 5 carbon (12 g/mol), 1 nitrogen (14 g/mol) and 10 hydrogen (1g/mol) atoms. Adding the molecular weights of the constituents atoms, we find that on average, approximately 10% of the protein weight is nitrogen. So, approximately $5\% \times 10\% = 0.5\%$ of the biomass in a plant is composed of nitrogen. Then, we can estimate that:

$$\begin{aligned}\text{Nitrogen per biomass per square meter} &= 0.5\% \times 10 \text{ kg/m}^2 \\ &= 0.05 \text{ kg/m}^2 \end{aligned} \tag{4}$$

Finally, to calculate fertilizer usage, we will assume that the fertilizer is completely composed of nitrogen for the simplicity of calculation. Then, for the entirety of the Central Valley, we need:

$$0.05 \text{ kg fertilizer/m}^2 \times 10^{10} \text{ m}^2 = 5 \times 10^8 \text{ kg fertilizer}. \tag{5}$$

(c) To estimate the amount of pesticide used every year, we will start from an easier estimation by thinking of how pesticide is sprayed using crop dusters. Crop duster is a small agricultural aircraft that can spray the pesticide while flying. We can assume that a typical crop duster can carry around 1 m^3 = 1000L of pesticide and cover an area of 1km × 1km per flight. Then, we can estimate the amount of pesticide used per square meter per year as

$$\text{Pesticide needed every year} = \frac{1000\,\text{L/year}}{1\,\text{km} \cdot 1\,\text{km}} = 10^{-3}\ \text{L/m}^2 \cdot \text{year} \tag{6}$$

For the entirety of the Central Valley, we need:

$$10^{10}\ \text{m}^2 \times 10^{-3}\ \text{L/m}^2 \cdot \text{year} = 10^7\ \text{L/year} \tag{7}$$

Assuming that the density of pesticide is the same as water ($\rho = 1$ kg/L), this leads to about 5×10^7 kg of pesticide used every year.

(d) For water usage, based on data from Figure 8 of https://fas.org/sgp/crs/misc/R44093.pdf, we can find that the total agricultural water used in Central Valley is around 25 million acre feet which is around $3{\times}10^{10}$ m^3 = $3{\times}10^{13}$ L, which is similar to our estimation.

For nitrogen fertilizer usage, based on data from Figure 1 of http://groundwaternitrate.ucdavis.edu/files/173452.pdf, we can estimate the total nitrogen usage in Central Valley as $\sim$ 400 Gigagram which is 4×10^8 kg – very close to our estimation of 5×10^8 kg.

Lastly, for pesticide usage, based on Agricultural Pesticide Mapping Tool we know that average pesticide usage is about 2.5 lbs/acre. Thus, the estimation of total pesticide usage is around 2.5 lbs/acre × 0.45 kg/lbs × 0.00025 acre/m^2 × 10^{10} m^2= 2.8×10^6 kg which is about twice less than our estimation.

2. To build a cell

Do problems 2.5 (ingredients in minimal media) and 3.7 (sugar budget of a cell) from PBOC2. Together, these two problems are intended to get you thinking about the wondrous process whereby a clear liquid with simple chemical ingredients is converted into biomass as shown in Figure. Amazing! In addition to working out the two problems given above, also work out an estimate for the volume of the headspace you see in Figure 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 have done 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 2: 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.

PBoC2 problem 2.5 – Ingredients in minimal media

(a) A standard *E. coli* cell is composed of approximately 3×10^6 proteins, 4×10^6 base pairs, and 2×10^7 lipids. These numbers are consistent with the numbers given in the chapter as well as those found in Table 1 of *Physiology of the Bacterial Cell* by Neidhardt, Ingraham and Schaecter. To determine the number of sugars needed to make a bacterium, we need to know how many carbon atoms are in a typical protein, a DNA base pair, and a standard

lipid. For this problem we will say that, on average, each amino acid has 5 carbon atoms, each sugar + base pair has 20 carbon atoms, and each lipid has 40 carbon atoms. Of course, these are all crude estimates and as with the entirety of the solution for this problem, should be seen as a simple estimate to give a feeling for the numbers. Given these numbers, the amount of carbon in each type of molecule is:

$$\begin{aligned} 3 \times 10^6 \,\text{proteins} \cdot \frac{300\,\text{amino acids}}{\text{protein}} \cdot \frac{5\,\text{carbons}}{\text{amino acid}} &= 4.5 \times 10^9\,\text{carbon atoms} \\ 4 \times 10^6\,\text{base pairs} \cdot \frac{20\,\text{carbons}}{\text{base pair}} &= 8 \times 10^7\,\text{carbon atoms} \\ 2 \times 10^7\,\text{lipids} \cdot \frac{40\,\text{carbons}}{\text{phospholipid}} &= 8 \times 10^8\,\text{carbon atoms.} \end{aligned}$$

We see that most of the carbon of a cell is invested in its proteins and we can neglect the contributions from DNA and lipids. Then we have about 5×10^9 carbon atoms in an *E. coli* cell and since one glucose molecule is made up of 6 carbons we can calculate the minimum number of glucose molecules that must be metabolized to make a single *E. coli* cell:

$$5 \times 10^9\,\text{carbons} \cdot \frac{1\,\text{sugar}}{6\,\text{carbons}} \approx 10^9\,\text{sugar molecules.}$$

In the case of nitrogen, we know we have about two nitrogen atoms per amino acid. We also know that each base pair has about eight nitrogen atoms. Finally, phospholipid molecules usually have one nitrogen atom. As a result, the number of nitrogen atoms corresponding to each macromolecule is

$$\begin{aligned} 3 \times 10^6\,\text{proteins} \cdot \frac{300\,\text{amino acids}}{\text{protein}} \cdot \frac{2\,\text{nitrogen atoms}}{\text{amino acid}} &= 2 \times 10^9\,\text{nitrogen atoms} \\ 4 \times 10^6\,\text{base pairs} \cdot \frac{8\,\text{nitrogen atoms}}{\text{base pair}} &= 3 \times 10^7\,\text{nitrogen atoms} \\ 2 \times 10^7\,\text{lipids} \cdot \frac{1\,\text{nitrogen atoms}}{\text{phospholipid}} &= 2 \times 10^7\,\text{nitrogen atoms.} \end{aligned}$$

As a result, the number of nitrogen atoms in the macromolecules of the bacterial cell is about 2×10^9.

Finally, each DNA base pair has two phosphate atoms, while each lipid molecule has one. As a result, the amount of phosphate in the cell's macromolecules is

$$\begin{aligned} 4 \times 10^6\,\text{base pairs} \cdot \frac{2\,\text{phosphate atoms}}{\text{base pair}} &= 8 \times 10^6\,\text{nitrogen atoms} \\ 2 \times 10^7\,\text{lipids} \cdot \frac{1\,\text{phosphate atom}}{\text{phospholipid}} &= 2 \times 10^7\,\text{nitrogen atoms} \end{aligned}$$

which means that the total amount of phosphate in macromolecules is about 3×10^7.

(b) As noted in the statement of the problem, the glucose is present in the medium at a concentration of 0.5g/100 mL. This implies that in 5 mL of minimal media, there are about

10^{-2} g of glucose. How many sugar molecules is this? Since the formula for glucose is $C_6H_{12}O_6$, the molecular mass is 180 Da. Hence, the number of sugars is

$$\#\text{ sugars} \approx \frac{10^{-2}\text{ g}}{180\ g/6\times 10^{23}\text{ molecules}} \approx 3\times 10^{19}\text{ glucose molecules.} \tag{8}$$

According to our estimate from part (a) (flawed though it is because it emphasizes only the construction material cost of making a cell and ignores the energetic requirements), it takes 10^9 sugar molecules to make a bacterium and hence our 5 mL culture can support roughly 10^{10} bacteria. This is consistent with our intuition because a saturated culture has roughly 10^9 cells/mL.

The molecular mass of NH_4Cl is about 50 Da. This means that in 5 mL of culture we have 6×10^{19} nitrogen atoms. As a result, if nitrogen was the limiting building block our culture would be able to support 3×10^{10} cell. From our estimations, when our culture starts running out of its carbon source it will also start running out of its nitrogen source.

Just to reiterate, the point of this estimate is to get a sense of the cellular inventory and how it relates to the molecular contents of the growth medium that is used for those cells. The precise numbers should not be taken too seriously.

PBoC2 problem 3.7 – The sugar budget in minimal medium
To do this problem, we will use the number of carbon atoms needed to synthesize a bacteria calculated above. Since a single glucose molecule contains 6 carbon atoms, the total number of carbon atoms needed to make a bacterial cell becomes

$$\frac{6\text{ carbon atoms}}{\text{glucose}}\times 10^9\,\frac{\text{glucose molecules}}{\text{cell}} = 6\ \times 10^9\frac{\text{carbon atoms}}{\text{cell}}. \tag{9}$$

The required rate of carbon intake therefore can be calculated as

$$\frac{6\times 10^9\,\text{carbon atoms}}{1200\,\text{s}} = 5\times 10^6\,\frac{\text{carbon atoms}}{\text{s}}. \tag{10}$$

As an aside, there are around 1,000 transmembrane proteins whose function is to import sugar. So each of these proteins must bring in 5,000 carbons (or, $\sim 1,000$ sugars) per second!

Volume of headspace needed for bacterial growth
To get the headspace volume required for the bacterial cells to grow, we will use the "divide and conquer" strategy. First, we calculate the total number of oxygen molecules needed (N_{O_2}) as the product of the number of bacterial cells available in the culture (N_{cells}) and the number of oxygen molecules required for building a single cell ($N_{O_2}^{\text{cell}}$), that is,

$$N_{O_2} = N_{\text{cells}}\times N_{O_2}^{\text{cell}}. \tag{11}$$

As we calculated above, the (maximum) number of bacterial cells in the 5ml of culture media is roughly $N_{\text{cells}}\approx 10^{10}$. The number of oxygen molecules needed for building a bacterial cell

can be calculated by using the amount of glucose molecules needed to make a cell ($\sim 10^9$) and the fact that 6 O_2 molecules are consumed for every glucose molecule, i.e. $N_{O_2}^{\text{cell}} = 6\times 10^9$.

Multiplying these two results, we find

$$N_{O_2} \approx 10^{10} \text{ bacteria} \cdot 6 \times 10^9 \ \frac{\text{molecules}}{\text{bacterium}} = 6 \times 10^{19} \,\text{molecules}. \tag{12}$$

Now, we need to convert the number of oxygen molecules to the volume of oxygen gas. We will use the fact that 1 mole of gas under standard temperature and pressure has a volume of 22.4 L. The volume of the oxygen gas can then be estimates as

$$\begin{aligned} V_{O_2} &= N_{O_2} \cdot \frac{22.4\,\text{L/mol}}{6 \times 10^{23}\,\text{molecules / mol}} = 6 \times 10^{19} \cdot \frac{22.4\,\text{L}}{6 \times 10^{23}} \\ &\approx 2 \times 10^{-3}\,\text{L} = 2\,\text{mL}. \end{aligned} \tag{13}$$

Since oxygen is roughly 20% of the air composition, we need to multiply by 5 to get the volume of air needed to support the growth of the bacteria culture, that is,

$$V_{\text{air}} = 5V_{O_2} = 10\,\text{mL}. \tag{14}$$

In Figure 2, we can see that typically we put 5 ml of media in 15 ml capacity tubes, which leaves a headspace of 10ml. In reality, the amount of oxygen needed for the bacterial culture would be more than what we have estimated, since we didn't take into account of the carbon/oxygen needed to synthesize the bacteria, and only considered the demands of providing the bacterial constituents. Thus, in reality, we have to shake the tubes so that air can circulate and be refilled for an overnight culture.

3. 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. Then, 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?

(b) 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.

(a) This questions is purposely somewhat open-ended to encourage you to explore and research about ATP synthase. Here are some key points we wanted you to uncover:

- The enzyme is essentially a motor. It extracts energy from the proton gradient across the membrane by allowing protons to flow through it (down the gradient), driving the rotation of some of its subunits.
- Stoichiometry of synthesis is 3 to 1, that is, 3 ATPs are synthesized per complete rotation of the enzyme. This is related to the 3-fold symmetry of the alpha and beta subunits which contain the catalytic sites.
- The rotation rate is about 6,000 rmp, or 100 Hz. So a single enzyme can synthesize around 300 ATPs per second.
- The enzyme's function is reversible. The energy to synthesize the new ATPs comes from the proton gradient. The enzyme is just a "messenger" of the proton motive force. If the proton gradient is reversed, the enzyme will hydrolyze ATP and pump protons "uphill," against the gradient.

One approach to estimate the ATPs needed to make a cell could be counting the molecules that make up a cell and then calculating the energetic cost of making each of these molecules. Here it makes sense to consider only the most abundant and/or costly molecules. These are the proteins, 2.4×10^6 of them per cell. What is the ATP cost of making a single protein? The polymerization of an single amino acid onto a polypeptide requires 4 ATPs. Therefore the total number of ATPs needed to synthesize the proteins of a cell becomes

$$2.4 \times 10^6 \,\text{proteins} \cdot \frac{300\,\text{amino acids}}{\text{protein}} \cdot \frac{4\,\text{ATP}}{\text{amino acid}} \sim 3 \times 10^9 \frac{\text{ATP}}{\text{cell}}. \tag{15}$$

All this ATP has to be synthesized in 20 min (doubling time), and thus, we calculate the required rate of ATP synthesis as

$$3 \times 10^9 \frac{\text{ATP}}{20 \text{ min}} \cdot \frac{1}{60} \frac{\text{min}}{\text{s}} \sim 2.5 \times 10^6 \frac{\text{ATP}}{\text{s}}. \tag{16}$$

Since the enzymatic rate of ATP synthesis is ~ 300/s, we find that roughly 10^4 ATP synthases are needed to meet the ATP production demands of the cell. In fact, 10^4 is quite close to the ribosomal profiling measurements (by Gene Wei Li in the spreadsheet) in MOPS complete media (which should give a doubling time close to our 1000 s estimate). They count very nearly 10^4 of each a, gamma, delta, and epsilon subunits, and roughly 3×10^4 for the alpha and beta subunits, as expected from the stoichiometry. There is some excess in the b subunit, and in the Schmidt *et al.* data, there is significant deviation between the expected and measured stoichiometry in all subunits. It is interesting to wonder if this is simply experimental error or a real effect of the cell failing to match the expected stoichiometry, or a combination of both. Nevertheless, the Schmidt, *et al.* data comes in very roughly at the 10^4 total functional ATP synthases per cell, in line with our estimate.

(b) In lecture, we argued that the dry mass of an *E. coli* cell is about 300 fg (the remaining 70% of the 1 pg being "wet"), and of that, about half or 150 fg is protein. The squares quite well with Figure 2b from Schmidt *et al.*, which shows that the total protein mass is between 100 and 200 fg depending on the growth condition and corresponding growth rate.

4. 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. To that end, to get a feeling for transcription, do problem 3.4 of PBoC2.

(a) The dry weight of an *E. coli* cell is roughly 30% of its total mass, half of which is protein. The total mass of the cell is estimated to be a picogram from the assumption that its density is nearly that of water and its volume roughly 1 μm^3. It is given in the problem that the β and β' subunits together have a mass of 300 kDa and comprise 0.5% of the protein mass. The number of β, β' subunit pairs is assumed to equal the number of RNA polymerases (RNAP). Putting all this together yields

$$\#\text{ of RNAP} = \frac{0.15 \times 10^{-12}\text{ g} \times 0.005}{3 \times 10^5\text{ Da} \times 1.6 \times 10^{-24}\text{ g/Da}} \approx 1.5 \times 10^3. \tag{17}$$

(b) Comparing Figs. (A) and (B), one sees that 40 seconds after rifampin addition roughly 1.5 kb of the DNA from the start site has become free of RNAP. The micrographs are aligned well enough that one can assume the left edge in all of them is the start site. Assuming that the last RNAP to initiate transcription did so at nearly the same time as rifampin addition, one can infer that this RNAP transcribed 1.5 kb of DNA in 40 seconds, implying an elongation rate of roughly 0.04 kb/sec. Making the same comparison of Figs. (A) and (C), indicates an elongation rate of 3.5 kb/70 sec = 0.05 kb/sec, or roughly 50 nucleotides/sec.

(c) The operon is roughly 6 kb long. Given the elongation rate in (b), one RNAP would require 6/0.05 = 120 seconds to complete a transcript. To estimate the rate at which transcripts of the operon are made, one needs the number of RNAP on the operon at any one time. Looking at the micrograph, one can make a rough count that under normal conditions there are 10 – 20 RNAP per kilobase and that the operon is roughly 6 kb long. This implies roughly $6 \cdot 15 = 90$ RNAP on the operon, and if each RNAP requires $\frac{6\text{ kb}}{0.05\text{ kb/sec}} = 120$ sec to complete transcription, then the transcript production rate is $120/90 \approx 1$ transcript per second. Alternatively, we can notice that the mean spacing between RNAPs is roughly 100 bp. The average speed of each such RNAP (from the previous part of the problem) is 50 nt/s. Hence, approximately every 2 sec another RNAP initiates transcription.

(d) During the course of a cell cycle, the cell must double the numbers of all its parts. Therefore if there are 20,000 ribosomes and the division time is 3000 seconds, the cell must produce $20,000/3000 = 6.7$ ribosomes per second, at least. This operon codes for the 16S and 23S ribosomal subunits, which are made directly from the RNA transcript, and each ribosome contains only one of each of these. This means that the rate of ribosome production is at most equal to the rate of transcription of this operon times the number of copies of this operon. From part (c) we know that it would require $6\,\text{kb} \times 20/\text{kb} = 120$ transcribing RNAP

to produce roughly one transcript per second. Therefore to make 6.7 transcripts per second requires $6.7 \cdot 120 \approx 800$ RNAP to be always transcribing the operon. Given the estimate in part (a), this is roughly half of all the RNAP in the cell. Further, this clearly points to the existence of more than one copy of the rRNA operon. Indeed, as the reader can see by consulting the rrnDB (the Ribosomal RNA Operon Copy Number Database), the average number of copies of this operon in *E. coli* is roughly seven.

5. 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 with *E. coli* growing with a division time of 24 minutes having 72,000 ribosomes per cell and slow growing *E. coli* with a division time of 100 minutes having a factor of ten fewer ribosomes with a count of $\approx$ 6800 ribosomes. In this problem, we will use our street fighting skills to explore the ribosomal density in another organism as shown in Figure 3, 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 above.

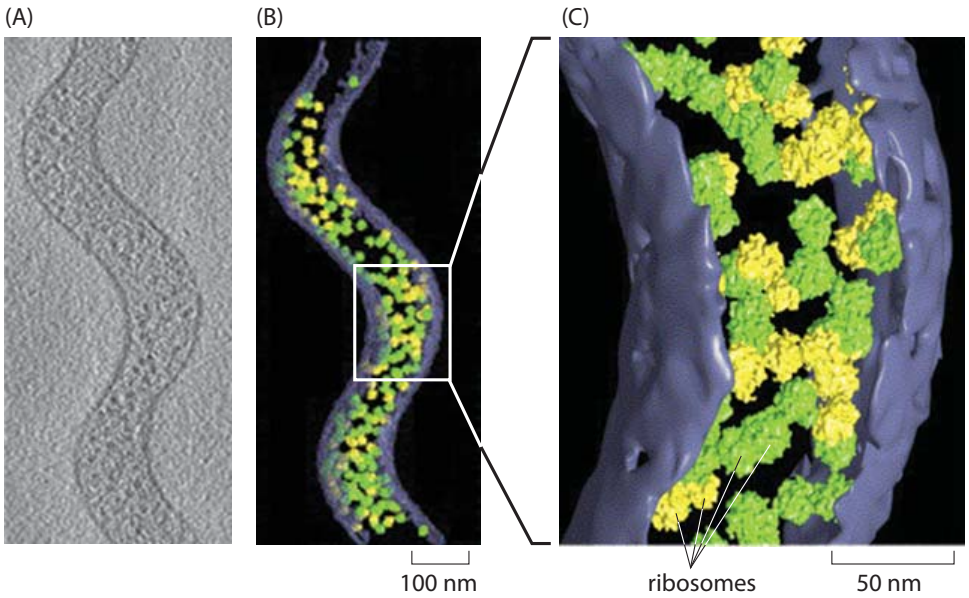

Figure 3: Satellite image of California's Central Valley.

In the close up view of the 3D reconstruction (panel C of the Figure) we can count 25 ribosomes labeled in green (high fidelity) and 17 ribosomes labeled in yellow (intermediate fidelity). Including 10 of the intermediate-fidelity ribosomes into our counting, we can say with high confidence that there are $N_{\text{close up}} \approx 35$ ribosomes in panel C.

Next, to estimate the volume of the cell section in panel C, we approximate it as a cylinder with a diameter of 100 nm and height of 200 nm, whose volume is given by

$$\begin{aligned} V_{\text{close up}} &\approx \frac{\pi \times (100\,\text{nm})^2}{4} \times 200\,\text{nm} \\ &\approx 2 \times 10^6\,\text{nm}^3. \end{aligned} \tag{18}$$

The estimated concentration of ribosomes in *Spiroplasma melliferum* then becomes

$$\begin{aligned}
\rho &= \frac{N_{\text{close up}}}{V_{\text{close up}}} \\
&= \frac{35}{2 \times 10^6\,\text{nm}^3} \\
&\approx 2 \times 10^{-5}\,\text{nm}^{-3} \\
&= 2 \times 10^{-5}\,\text{nm}^{-3} \times \left(\frac{10^3\,\text{nm}}{1\,\mu\text{m}}\right)^3 \\
&= 2 \times 10^4\,\mu\text{m}^{-3}. \qquad (19)
\end{aligned}$$

Our estimate of 20,000 ribosomes per μm^3 falls nicely within the range observed for *E. coli* cells, which have a volume of $\sim 1\,\mu\text{m}^3$ and hence, ribosome density range of $\sim 7,000 - 70,000$ per μm^3.

6. RNA vs. Protein

Using the kind of estimates we have talked about in class, give a simple characterization of the relative sizes of mRNAs and the proteins they code for. Specifically, first comment on the mean mass of amino acids and nucleotides as well as their typical physical sizes. Use both of these metrics as a way to provide a rough sense of the relative sizes both in mass and physical dimensions of proteins and the mRNAs that code for them.

Looking at the atomic composition of amino acids, we can say that on average they contain two oxygen (16 g/mol), 5 carbon (12 g/mol), 1 nitrogen (14 g/mol) and 10 hydrogen (1 g/mol) atoms. Adding the weights of the constituents atoms, we find that the mass of a typical amino acid is

$$\begin{aligned} m_{\rm aa} &\approx (2\times 16 + 5\times 12 + 14 + 10)\ \frac{\rm g}{\rm mol} \\ &\approx 120\frac{\rm g}{\rm mol} \approx 100\frac{\rm g}{\rm mol} \\ &= 100\,{\rm Da}. \end{aligned} \tag{20}$$

To estimate the size of a typical amino acid, we use $1\,{\rm g/cm^3}$ as a street fighter's estimate for its average density, and, modeling it as a cube of side length $\ell_{\rm aa}$, obtain

$$\begin{aligned} \ell_{\rm aa}^3 &= \frac{m_{\rm aa}}{1\,{\rm g/cm^3}} \\ &= 100\,{\rm Da}\times\frac{1\,{\rm g}}{6\times 10^{23}\,{\rm Da}}\times 1\frac{\rm cm^3}{\rm g}\times\left(\frac{10^7\,{\rm nm}}{1\,{\rm cm}}\right)^3 \\ &\approx 0.2\,{\rm nm}^3 \Rightarrow \qquad (21) \\ \ell_{\rm aa} &\approx 0.5\,{\rm nm}. \qquad (22) \end{aligned}$$

Finally, a typical protein is made out of roughly 300 amino acids. This suggests that the typical protein mass, volume and radius (assuming tight packing of amino acids) are, respectively,

$$m_{\rm protein} \approx 300\,{\rm aa}\times 100\,\frac{\rm Da}{\rm aa} = 30\,{\rm kDa}, \tag{23}$$

$$V_{\rm protein} \approx 300\,{\rm aa}\times 0.2\,\frac{\rm nm^3}{\rm aa} = 60\,{\rm nm}^3, \tag{24}$$

$$\frac{4}{3}\pi R_{\rm protein}^3 = V_{\rm protein} \Rightarrow R_{\rm protein} \approx \left(\frac{V_{\rm protein}}{4}\right)^{1/3} = (15\,{\rm nm}^3)^{1/3} \approx 2.5\,{\rm nm}. \tag{25}$$

Let us now do similar estimates for nucleotides. We can say that, on average, a nucleotide contains 1 oxygen (16 g/mol), 5 carbon (12 g/mol), 3 nitrogen (14 g/mol) and 5 hydrogen (5 g/mol) atoms. In addition, the sugar phosphate backbone has 4 carbon (12 g/mol), 1

phosphate (96 g/mol) and 5 oxygen (16 g/mol) atoms. Adding all the weight together, we estimate the contribution of a single nucleotide to the mass of an mRNA to be

$$\begin{aligned} m_{\text{nucleotide}} &= (16 + 5 \times 12 + 3 \times 14 + 5)\,\frac{\text{g}}{\text{mol}} + (4 \times 12 + 96 + 5 \times 16)\,\frac{\text{g}}{\text{mol}} \\ &\approx (120 + 220)\,\frac{\text{g}}{\text{mol}} \\ &\approx 350\,\text{Da}. \end{aligned} \tag{26}$$

Since three nucleotides are needed to code a single amino acid, a typical mRNA would have contain ~ 1000 nucleotides, and thus, the weight of a typical mRNA will be

$$\begin{aligned} m_{\text{mRNA}} &\approx 350\,\frac{\text{Da}}{\text{nucleotide}} \times 1000\,\text{nucleotides} \\ &= 350\,\text{kDA} \\ &\approx 10 \times m_{\text{protein}}. \end{aligned} \tag{27}$$

Lastly, the length of a typical mRNA is $1000 \times 1/3\,\text{nm} \approx 300\,\text{nm}$. Unlike proteins, mRNAs do not having a tightly folded structure. To estimate the size of the region that they occupy in the cell, we can imagine that each 20 nucleotide-long mRNA chunk has an independent orientation (i.e., the persistence length is ~ 7 nm). Therefore, an mRNA could be approximates as a chain of 50 connected chunks with random orientations, which occupy a region in space that a roughly size of $7\,\text{nm} \times \sqrt{50} \approx 50\,\text{nm}$, which is an order of magnitude higher than the size of a typical protein.