BE/APh161: Physical Biology of the Cell
Homework 1
Due Date: Wednesday, January 10, 2024

Solutions by: Nicholas Gravina, Tom Röschinger, Jordan Santana

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 × 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 the parts of chap. 2, "Setting the Scales of Living Things" of the upcoming third edition of Physical Biology of the Cell you will find with the online link to this homework. 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. 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

(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.

Solution: We can get to the thickness of the oil layer by simply dividing the volume of oil by the area it spread across. For the volume, we will take "not more than a teaspoon" to be a few mL or few cm^3 and the "perhaps half an acre" to be a few thousand m^2. The rest of the work involves unit conversions to get our answer into meaningful units:

$$\text{height} = \frac{\text{volume}}{\text{area}} = \frac{\text{few cm}^3}{\text{few} \times 10^3\ \text{m}^2} \times \frac{\text{m}^2}{10^4\ \text{cm}^2} = 10^{-7}\ \text{cm} \times \frac{10^7\ \text{nm}}{\text{cm}} = 1\ \text{nm}. \tag{1}$$

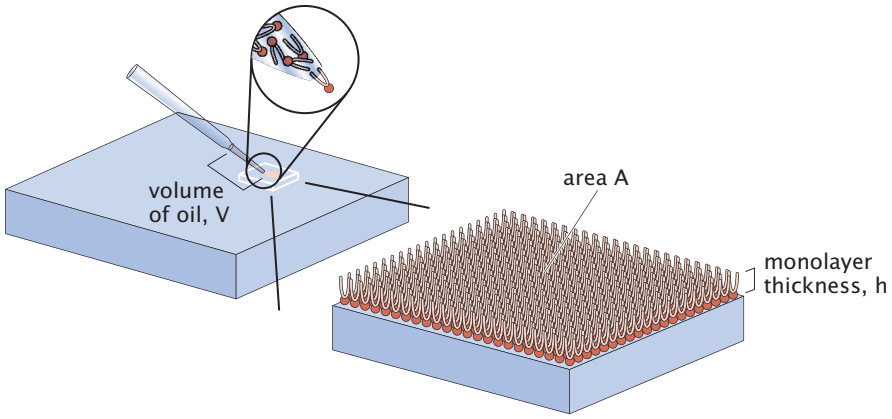

Figure 1: Putting oil on water to measure molecular dimensions. Here we see that the lipid molecules form a monolayer.

(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?

Solution: Assuming the oil is roughly the density of water, which is not unreasonable for an order-of-magnitude estimate, our few mL of oil correspond to a few g of oil. Using the provided molecular mass and Avogadro's number, we arrive at a total number of lipid molecules:

$$\text{few g} \times \frac{\text{mol}}{1000\,\text{g}} \times \frac{6 \times 10^{23}\,\text{lipid molecules}}{\text{mol}} \approx 2 \times 10^{21}\,\text{lipids}. \tag{2}$$

To get the area of the head of lipid, we simply need to divide the area the oil was spread across by this number of lipids:

$$\text{area} = \frac{\text{few} \times 10^3\,\text{m}^2}{2 \times 10^{21}\,\text{lipids}} \approx \frac{10^{-18}\,\text{m}^2}{\text{lipid}} \times \frac{10^{18}\,\text{nm}^2}{\text{m}^2} = 1\,\text{nm}^2/\text{lipid}. \tag{3}$$

Comparing our results to the known values, we find that the lipid *bi*layer is around 4 nm thick (BNID:105298), meaning that each lipid is about 2 nm, which is only a factor of two off from our estimate in part (a). Pretty good considering how crude the experiment was and how imprecise the descriptions are! For the surface area of a lipid, we see that value is 0.5 nm^2 (BNID:106993), which is again just a factor of two off from our estimate.

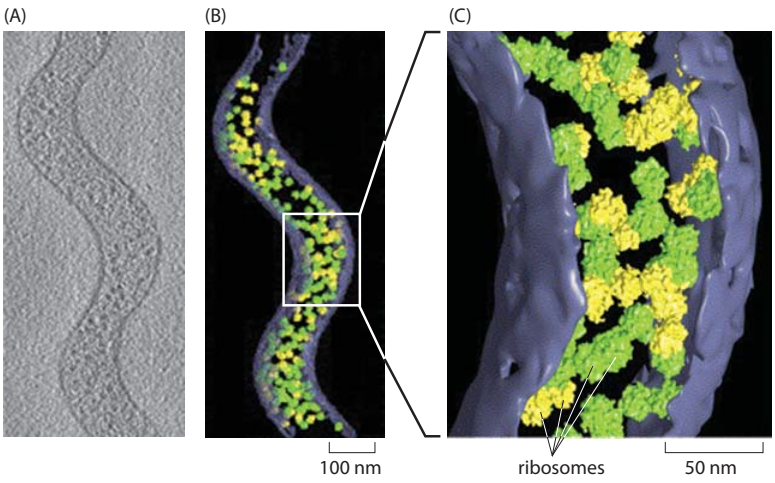

Figure 2: Cryo EM study of a bacterial cell. These images are of the tiny bacterium, *Spiroplasma melliferum*. Using algorithms for pattern recognition and classification, components of the cell such as ribosomes were localized and counted. (A) Single cryo-electron microscopy image. (B) 3D reconstruction 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).

3. Street fighting the ribosome

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 ten fewer ribosomes with a count of ≈ 6800 ribosomes.

(a) In this part of the 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.

Solution: In the close up view of the 3D reconstruction (panel C of the Fig 2) 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

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

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

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

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.

(b) In a beautiful turn of the millennium paper by Tania Baker and Stephen Bell whose abstract is shown in Figure 3, they imagined a world in which DNA polymerase was the size of a FedEx truck and explored what copying DNA would look like. Write a one-paragraph abstract of your own which carries out a similar analysis, but this time for the ribosome.

Cell, Vol. 92, 295–305, February 6, 1998, Copyright ©1998 by Cell Press

Polymerases and the Replisome: Machines within Machines

Review

Tania A. Baker* and Stephen P. Bell
Department of Biology
*Howard Hughes Medical Institute
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Synthesis of all genomic DNA involves the highly coordinated action of multiple polypeptides. These proteins assemble two new DNA chains at a remarkable pace, approaching 1000 nucleotides (nt) per second in *E. coli*. If the DNA duplex were 1 m in diameter, then the following statements would roughly describe *E. coli* replication. The fork would move at approximately 600 km/hr (375 mph), and the replication machinery would be about the size of a FedEx delivery truck. Replicating the *E. coli* genome would be a 40 min, 400 km (250 mile) trip for two such machines, which would, on average make an error only once every 170 km (106 miles). The mechanical prowess of this complex is even more impressive given that it synthesizes two chains simultaneously as it moves. Although one strand is synthesized in the same direction as the fork is moving, the other chain (the lagging strand) is synthesized in a piecemeal fashion (as Okazaki fragments) and in the opposite direction of overall fork movement. As a result, about once a second one delivery person (i.e., polymerase active site) associated with the truck must take a detour, coming off and then rejoining its template DNA strand, to synthesize the 0.2 km (0.13 mile) fragments.

In this review we describe our current understanding of the organization and function of the proteins of the replication fork and how these complexes are assembled at origins of replication. Understanding the architecture of DNA polymerases is relevant to RNA polymerases as well, as the core of the polynucleotide polymerization machine appears to be similar for all such enzymes. In the discussion of the replisome, we particularly focus on features shared by the machinery from different organisms.

Polymerases: Template-Directed Phosphoryl Transfer Machines

Synthesis of the new DNA strands occurs as a result of a collaboration between the synthetic capacities of multiple polymerases. Two types of polymerases are required: primases, which start chains, and replicative polymerases, which synthesize the majority of the DNA (Kornberg and Baker, 1992). The replication fork, however, contains at least three distinct polymerase activities: a primase and a replicative polymerase for each of the two template strands. In *E. coli*, primase is a single polypeptide, and the replicative polymerase is a dimer of DNA polymerase (pol) III core and several accessory proteins that together form the pol III holoenzyme (reviewed in Marians, 1992; Kelman and O'Donnell, 1995). Similarly, phage T4 has one primase and one replicative polymerase that appears to function as a dimer (Alberts, 1987; Munn and Alberts, 1991). The situation in eukaryotic cells is slightly different (Stillman, 1994). The primase is in a tight complex with a DNA polymerase (pol α) and eukaryotic cells have two distinct replicative polymerases: polymerase δ (pol δ) and polymerase ϵ (pol ϵ).

All the replicative polymerases have one large subunit that contains the polymerase active site and, with the exception of pol α–primase, the same subunit or an associated polypeptide carries a proofreading $3' \rightarrow 5'$ exonuclease. The polymerase subunits also interact with proteins that dramatically influence their association with DNA. In *E. coli*, the replicative polymerase is found in a complex with proteins that control polymerase processivity; this holoenzyme, consists of 10 distinct polypeptides (Kelman and O'Donnell, 1995). In contrast, neither the T4 nor the eukaryotic polymerases copurify in a complex with the processivity factors (Alberts, 1987; Stillman, 1994). Therefore, these proteins are called accessory proteins rather than subunits (see Table 1).

Polymerase Architecture. The central feature of all the known polymerase structures is the existence of a large cleft comprised of three subdomains referred to as the fingers, palm, and thumb by virtue of the similarity of the structures to a half-opened right hand (Figure 1; polymerase structures are reviewed in Joyce and Steitz, 1994, 1995; Sousa, 1996). A diverse set of polymerases—

Figure 3: Abstract of a paper from Tania Baker where she maps the action of DNA polymerase onto human length scales to give a sense of its amazing properties. This parable is the basis of your own analysis of the ribosome. Adapted from Baker TA and Bell, SP Cell, Vol. 92, 295-305, February 6, (1998).

Solution: Relevant bionumbers:
• Peptite chain elongation (*E. coli*): 12-21 aa/s (BNID 100059).
• Misincorporation rate of amino acids in translation (*E. coli*): $10^{-4} - 10^{-3}$ (BNID 103454).
• Characteristic average size of mRNA: 1000 nucleotides (BNID 100022).
• Width of mRNA backbone: 2 nm (Cell Biology by the Numbers).
• Ribosome diameter (mammalian cells): 30 nm (BNID 100483).

Relevant scaled-up bionumbers assuming the same rule of thumb for length of mRNA per nucleotide as for length of DNA per base pair (1/3 nm) and a 1 m width for mRNA:
• Speed of ribosome along mRNA $\approx 8\ ms^{-1} \approx 30km/h$ $(20mph)$
• Size of ribosome $\approx 15\ m$
• Time to translate characteristic protein $\approx 20\ s$
• Length of characteristic mRNA $\approx 150\ m$
• Error rate for protein synthesis $\approx$ 1 per 500 - 5000 m (we will take the geometric mean $\approx 1500\ m$).

If mRNA were 1 m in width, then the following statements would roughly describe E. coli replication. Ribosomes would move along the mRNA at approximately 30 km/h (20 mph), and a ribosome would be roughly the size of a semi-truck trailer. Translating a typical strand of mRNA would be a 150 m, 20s trip with translational errors occuring roughly once every 1500 m (or roughly once every 10 translations).

4. 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*.

Solution: If we look at Fig. 2.1 in PBoC, we see that the shape of an E. coli bacterium can be approximated as a cylinder with two hemispheres on the two ends. Using the same figure, we can approximate the height of the cylinder as 1 μm, and the diameter of the hemisphere as 0.5μm. Then, we can calculate the volume as

$$V = \pi(0.5\mu m)^2 \times (1\mu m) + 4/3\pi(0.5\mu m)^3 = 1.309\mu m^3$$

For the mass, we can assume that E.coli is roughly the same density as water, 1 g/cm^3. It then follows that the mass is approximately 1 pg.

(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.

Solution: Given the volume of approximately 1 $\mu\text{m}^3 = 10^{-15}$ L derived in part (a), it follows that the concentration of 1 molecule per *E. coli* cell should be

$$[1 \text{ molecule}/\textit{E. coli} \text{ cell}] \approx \frac{10^{15} \text{ molecules}}{L} \cdot \frac{\text{mol}}{6 \times 10^{23} \text{ molecules}} \approx 1 \text{ nM}. \tag{6}$$

(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.

Solution: The molecular composition given for a "typical" amino acid has a molecular mass of ≈ 100 Da. Since the total mass of all protein in the cell constitutes about 1/6 pg, we arrive at the total number of amino acids in a cell:

$$\frac{1}{6} \text{ pg} \cdot \frac{6 \times 10^{11} \text{ Da}}{\text{pg}} \cdot \frac{\text{aa}}{100 \text{ Da}} = 10^9 \text{ aa}. \tag{7}$$

Treating a "typical" protein as having 300 amino acids, these 10^9 amino acids correspond to few $\times 10^6$ proteins in our *E. coli* 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)?

Solution: If we model the bacterium as being tightly packed with proteins, we can estimate the number of proteins per cell by using protein volume. Assume the diameter of a typical (folded) protein is 5 nm $\approx f$ nm (from Bionumbers site) and that proteins are spherical. The volume of a typical sphere is then

$$V_{\text{protein}} = \frac{4}{3}\pi(f \text{ nm})^3 \approx 100 \text{ nm}^3. \tag{8}$$

Assuming the bacterium is tightly packed with proteins, we then estimate that there are about

$$\frac{1\mu m^3}{100 \text{ nm}^3} = 10^7 \text{ proteins per } \textit{E. coli}, \tag{9}$$

or more realistically slightly under given the efficiency of spherical tight packing. This result differs from the estimate in part (c) by roughly an order of magnitude, most likely arising from the assumption in (d) that protein packs the entire volume of the bacterium. Given that protein makes up a fraction of the dry mass, which is in turn a fraction of the cell mass, it is reasonable to expect the cell to not be tightly packed with protein throughout.

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

Solution: Turning now to the number of nucleotides, we find that the molar mass of the "typical" nucleotide is around 300 Da, giving us a total number of nucleotides

$$\frac{1}{6}\,\text{pg} \times \frac{6 \times 10^{11}\,\text{Da}}{\text{pg}} \times \frac{\text{nt}}{300\,\text{Da}} \approx 3 \times 10^{8}\,\text{nt}. \tag{10}$$

We see that the number of nucleotides that are found in the genome (around 5 million) accounts for a negligible fraction of the entire nucleotide composition.

(f) 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?

Solution: We now turn to the source of non-genomic nucleotides, namely rRNA found in ribosomes. First, let's estimate the number of ribosomes that would be needed to synthesize the few million proteins discussed earlier. If we consider fast-growing cells, we have around 1200 seconds to create the billion peptide bonds between amino acids. This gives us a global synthesis rate of

$$\frac{10^{9}\text{aa}}{1200\text{s}} \approx 800,000\text{aa/s} \tag{11}$$

With each ribosome churning away at 15 aa/s, we need a total of 800,000/15 or 50,000 ribosomes to accommodate this global rate of protein production. As a sanity check, we see that this number meshes well with those provided in the previous problem.

Now to address how much of our nucleotide mass it tied up in these ribosomes, we refer to the total mass of a ribosome, 2700 kDa (BNID:100118), of which about 60% is RNA. This means that each ribosome has roughly 1600 kDa of nucleotides. Returning to the typical nucleotide mass of 300 Da, we arrive at

$$\frac{1600\ \text{kDa}}{\text{ribosome}} \times \frac{\text{nt}}{0.3\ \text{kDa}} = 5,000\ \text{nt/ribosome} \tag{12}$$

Multiplying by the total number of ribosomes, we arrive at 2.5×10^{8} total nt found in all ribosomes, which we see accounts for the vast majority of the total nucleotide compostion we found earlier.

(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.

Solution: This problem has divided up the bacterium into 2/3 wet mass (which we assume to be simply water) and 1/3 dry mass, which we have simplified to be described solely by proteins and nucleic acids. We must now determine the mol ratio among these components comprising the *E. coli* cell.

Given the total mass 1 pg estimated in part (a), and using the molar mass 18 g/mol for water (H_2O), there are

$$\frac{2}{3} \cdot 10^{-12}\, g \times \frac{1\,\text{mol}}{18\,\text{g}} \approx 40 \times 10^{-15}\,\text{mol of wet mass (water)}. \tag{13}$$

From part (c), we know there are 10^9 amino acids making up a cell, so there are

$$10^9 \cdot \frac{\text{mol}}{6 \times 10^{23}} \approx 2 \times 10^{-15}\,\text{mol (aa)}. \tag{14}$$

Finally, in part (e) we determined there are approximately 3×10^8 total nucleotides in the bacterium, so we have

$$3 \times 10^8 \cdot \frac{\text{mol}}{6 \times 10^{23}} = 0.5 \times 10^{-15}\,\text{mol (nt)}. \tag{15}$$

We now conclude that an approximate formula for the stoichiometry of a bacterium is

$$H_2O :\ \text{aa} :\ \text{nt} \approx 40 : 2 : 0.5 \equiv 80 : 4 : 1. \tag{16}$$

5. 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*.

Solution: 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:

$$\begin{aligned} 3 \times 10^6 \text{ proteins} \cdot \frac{300 \text{ amino acids}}{\text{protein}} \cdot \frac{5 \text{ carbons}}{\text{amino acid}} &= 4.5 \times 10^9 \text{ carbon atoms} \\ 4 \times 10^6 \text{ base pairs} \cdot \frac{20 \text{ carbons}}{\text{base pair}} &= 8 \times 10^7 \text{ carbon atoms} \\ 2 \times 10^7 \text{ lipids} \cdot \frac{40 \text{ carbons}}{\text{phospholipid}} &= 8 \times 10^8 \text{ carbon atoms.} \end{aligned} \tag{17}$$

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. We then have about 5×10^9 carbon atoms in an *E. coli* cell.

(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*.

Solution Following the same approach as in part (a), we will say that, on average, each amino acid has 2 nitrogen atom, each sugar + base pair has 8 nitrogen atoms, and each lipid has 1 nitrogen atom. Based on these numbers, the amount of nitrogen in each type of molecule is:

$$\begin{aligned} 3 \times 10^6 \text{ proteins} \cdot \frac{300 \text{ amino acids}}{\text{protein}} \cdot \frac{2 \text{ nitrogens}}{\text{amino acid}} &= 1.8 \times 10^9 \text{ nitrogen atoms} \\ 4 \times 10^6 \text{ base pairs} \cdot \frac{8 \text{ nitrogens}}{\text{base pair}} &= 3.2 \times 10^7 \text{ nitrogen atoms} \\ 2 \times 10^7 \text{ lipids} \cdot \frac{1 \text{ nitrogen}}{\text{phospholipid}} &= 2 \times 10^7 \text{ nitrogen atoms.} \end{aligned} \tag{18}$$

Once again, we can largely neglect the contributions from DNA and lipids and conclude. There are then about 1.8×10^9 nitrogen atoms in an *E. coli* cell.

(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?

Solution: As noted in the statement of the problem, the glucose is present in the medium at a concentration of 0.2 g/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

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

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.

We may similarly analyze cell nitrogen consumption in a 5 mL culture using minimal media. As stated in the problem, minimal media requires ammonium chloride (formula NH_4Cl) present at a concentration 100 mM. Therefore, in 5 mL of minimal media, the number of ammonium chloride molecules present is

$$5\,\text{mL}\cdot\frac{\text{L}}{10^3\,\text{mL}}\cdot\frac{100\times 10^{-3}\,\text{mol}}{L}\cdot\frac{6.022\times 10^{23}}{\text{mol}} \approx 10^{20}\text{ ammonium chloride molecules.} \tag{20}$$

In part (b) we estimated that an *E. coli* cell is made up of about 1.8×10^9 nitrogen atoms, so by comparison with the result above, we again estimate that our 5 mL culture can support roughly 10^{10} bacteria, consistent with our approach analyzing glucose consumption.

(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.

Solution: Since a single glucose molecule contains 6 carbon atoms, we can simply use our result from part (a) to obtain a number of glucose molecules:

$$\frac{5\times 10^9\text{ carbons}}{\text{cell}}\times\frac{\text{glucose molecules}}{6\text{ carbons}} \approx \frac{10^9\text{ glucose molecules}}{\text{cell}}. \tag{21}$$

The required rate of carbon intake therefore can be calculated as

$$\frac{5\times 10^9\text{ carbon atoms}}{1200\text{ s}} \approx 4\times 10^6\frac{\text{carbon atoms}}{\text{s}}. \tag{22}$$

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!

(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.

Solution: 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,

$$N_{O_2} = N_{\text{cells}} \times N_{O_2}^{\text{cell}}. \tag{23}$$

As we calculated above, the (maximum) number of bacterial cells in the 5 mL 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

$$N_{O_2} \approx 10^{10} \text{ bacteria} \cdot 6 \times 10^9 \frac{\text{molecules}}{\text{bacterium}} = 6 \times 10^{19} \text{ molecules}. \tag{24}$$

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

$$V_{O_2} = N_{O_2} \cdot \frac{22.4 \text{ L/mol}}{6 \times 10^{23} \text{ molecules/mol}} = 6 \times 10^{19} \cdot \frac{22.4L}{6 \times 10^{23}} \approx 2 \times 10^{-3} L = 2 \text{ mL}. \tag{25}$$

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,

$$V_{\text{air}} = 5V_{O_2} = 10 \text{ mL}. \tag{26}$$

In Figure 4, we can see that typically we put 5 mL of media in 15 mL capacity tubes, which leaves a headspace of 10 mL. 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 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.

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.

6. 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.

Solution: We will assume that winter is too cold for the crops to grow (December-February). Due to the weather of California, there is limited rainfall in the area, so we will assume that all the water crops use to grow come 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 litres of water are needed everyday to irrigate the crops. 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

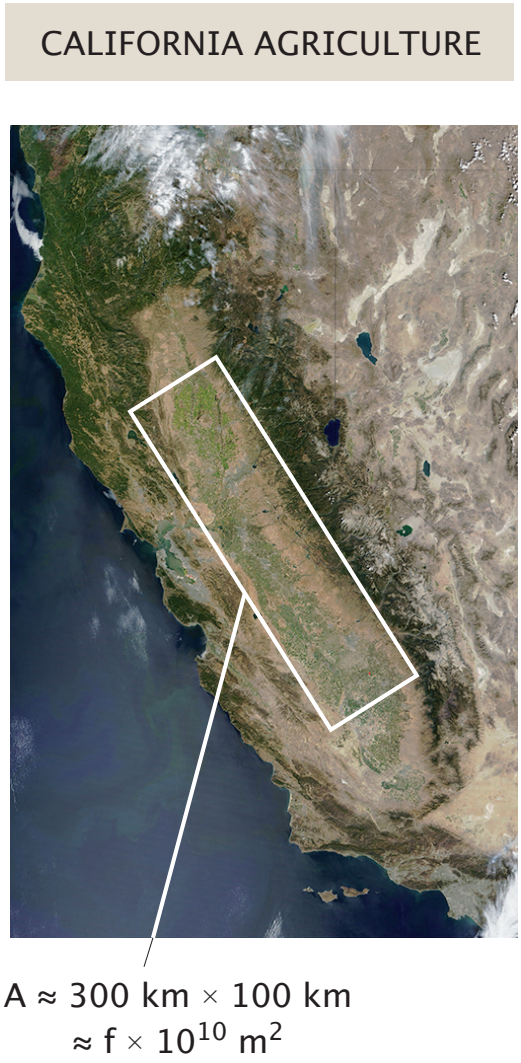

Figure 5: Satellite image of California's Central Valley.

the flower. That would give us

$$\text{water needed to irrigate crops in a unit area per day} = \frac{0.25 \text{ L/day}}{20 \text{ cm} \cdot 20 \text{ cm}} = \frac{0.25 \text{ L/day}}{0.04 \text{ m}^2} \approx 5 \text{ L/m}^2 \cdot \text{day}. \tag{27}$$

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

$$\underbrace{10^{10} \text{ m}^2}_{\text{Central Valley area}} \times 5 \,\text{L/m}^2 \cdot \text{day} \times 30 \text{ days/month} \times \underbrace{9 \text{ months/year}}_{\text{crop season}} \approx 10^{13} \text{L/year}. \tag{28}$$

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) 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.

Solution: 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 metre 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 ≈ 10 to get

$$\text{Biomass per square metre} \approx 10 \text{ kg/m}^2. \tag{29}$$

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 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

$$\text{Nitrogen per biomass per square metre} = 0.5\% \times 10 \text{ kg/m}^2 = 0.05 \text{ kg/m}^2 \tag{30}$$

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

$$0.05 \text{ kg fertilizer/m}^2 \times 10^{10} \text{ m}^2 = 5 \times 10^8 \text{ kg fertilizer.} \tag{31}$$

(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.

Solution: To estimate the pesticide used every year, we will start from an easier estimation by thinking of how pesticide is sprayed using crop dusters. A 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 $\text{m}^3 = 1000\text{L}$ of pesticide and cover an area of 1km×1km per flight. Then, we can estimate the amount of pesticide used per square metre per year:

$$\text{Pesticide needed every year} = \frac{1000\text{L/year}}{1\text{km} \cdot 1\text{km}} = 1 \times 10^{-3} \text{ L/m}^2 \cdot \text{year.} \tag{32}$$

For the entirety of the Central Valley, we need

$$10^{10} \text{ m}^2 \times 10^{-3} \text{ L/m}^2 \cdot \text{year} = 10^7 \text{ L/year.} \tag{33}$$

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

Solution: For water usage, based on data from Figure 8 of California Agricultural Production and Irrigated Water Use, we can calculate the total agricultural water used in Central Valley is around 25 million acre feet which is around 3×10^{10} $\text{m}^3 = 3 \times 10^{13}$ L which is similar to our estimation.

For nitrogen fertilizer usage, based on data from Figure 1 of Nitrogen Fertilizer Loading to Groundwater in the Central Valley, we can estimate that total nitrogen usage in Central Valley is about 400 Gigagram which is about 4×10^8 kg which is very close to our estimation of 5×10^8 kg.

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 slightly less than our estimation.

6. 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 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.

Solution: Using a radius of 50 nm and treating the virus as a sphere, we can estimate the volume of the virus as

$$\frac{4}{3}\pi(50\text{ nm})^3 \approx 5\times 10^5 \text{nm}^3 \times \frac{\mu\text{m}^3}{10^9\mu\text{m}^3} = 5\times 10^{-4}\mu\text{m}^3 \tag{34}$$

or 1/2000th the volume of an *E. coli* cell. Assuming a similar density to E. coli, we get that a SARS-CoV-2 virion is about 1/2000th of a pg, or 0.5 fg. We can find that the true mass of the viron is approximately 1 fg (DOI: 10.7554/eLife.57309), which is just within a factor of two of our estimate.

As for the mass of RNA, we will return to the roughly 300 Da nucleotide, as discussed in Problem 3. With a 30 kb genome, we get a total RNA mass of

$$300\ \frac{\text{Da}}{\text{bp}} \times 3\times 10^4\ \text{bp} = 9\times 10^6\ \text{Da} \times \frac{1.7\times 10^{-9}\ \text{fg}}{\text{Da}} \approx 0.015\ \text{fg} \tag{35}$$

This means that RNA makes up just $\approx 0.015/0.5 = 3\%$ of the mass of the SARS-CoV-2 based on our estimates. (I was expecting it to be higher given my general impression that viral genomes are tightly packed.)

(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?

Solution: While the RT-PCR assays gives us information about how many viral genomes (and thus the number of viruses) are found in an actively infected patient, the TCID50 assay address an entirely different question of how many viruses are required for causing an infection. So for estimating the mass of virions in a patient, we will use the assay that involves patient tissue. Given the relative size of the lungs to the rest of the body, I estimate them to weigh around 1 kg. Furthermore I will use the geometric mean of the low and high values that the assay yields and assume that 10^7 RNA copies per gram is the typical value. This gives us a total virion mass of

$$10^7 \frac{\text{virions}}{\text{g}} \times 1000 \text{ g} = 10^{10} \text{ virions} \times 0.5 \frac{\text{fg}}{\text{virion}} = 5 \times 10^9 \text{ fg} = 5\mu g \text{ per patient.} \tag{36}$$

For the fraction of cells that are infectious, we will refer to the TCID50 assay. Given the nature of viral infection once one cell has been successfully infected and lysed, releasing a new cohort of viruses, it is effectively game over. So we can think of infection as a Poisson process, by which each virus either "succeeds" at infection or "fails". Given that it takes around 1000 viruses to reliably cause infection, we can assume that around 1 in 1000 viruses are actually infectious, or 0.1%.

(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.

Solution: There have sadly been 100 million COVID cases. If we assume that our values from part (b) are representative of the typical case, we get a worldwide viral mass of

$$10^8 \text{ cases} \times 0.5 \frac{\mu\text{g}}{\text{case}} = 5 \times 10^8 \ \mu\text{g} = 500 \ g \tag{37}$$

It's wild to think how much havoc has been caused by a mere half a kilogram of biological mass. This estimate disregards viruses that are not in people, but given that the virus has no way to replicate on surfaces or outside of a living host, I would image that this would have a minor effect on the estimate.