BE/APh161: Physical Biology of the Cell
Homework 1
Due Date: Wednesday, January 14, 2026

Solutions by: Nicholas Gravina, Tom Röschinger, Jordan Santana, Sara Mahdavi

"The main obstacle to progress is not ignorance, but the illusion of knowledge." Ronald Graham in the Science Lives series of interviews by the Simons Foundation

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. 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). On the subject of AI, I use chatbots every single day and to great effect. That said, I would truly prefer that you do not use chatbots on this homework because these exercises are very helpful for the development of your own intuition.

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. Though we will routinely use AI in this course, this is NOT one of those moments. Here we are interested in you developing your most truly original and authentic thoughts.

2. William Harvey and the circulation of the blood

In 1628, William Harvey published one of the most profound scientific works of all time: *Exercitatio Anatomica de Motu Cordis et Sanguinis in Animalibus* (An Anatomical Disquisition on the Motion of the Heart and Blood in Animals). In this short treatise, Harvey shattered centuries of Galenic medical orthodoxy and demonstrated that blood does not ebb and flow in the body like tides, but circulates in a closed loop, pumped by the heart. His argument was not built on microscopes or chemical analysis. It was based on *estimation*.

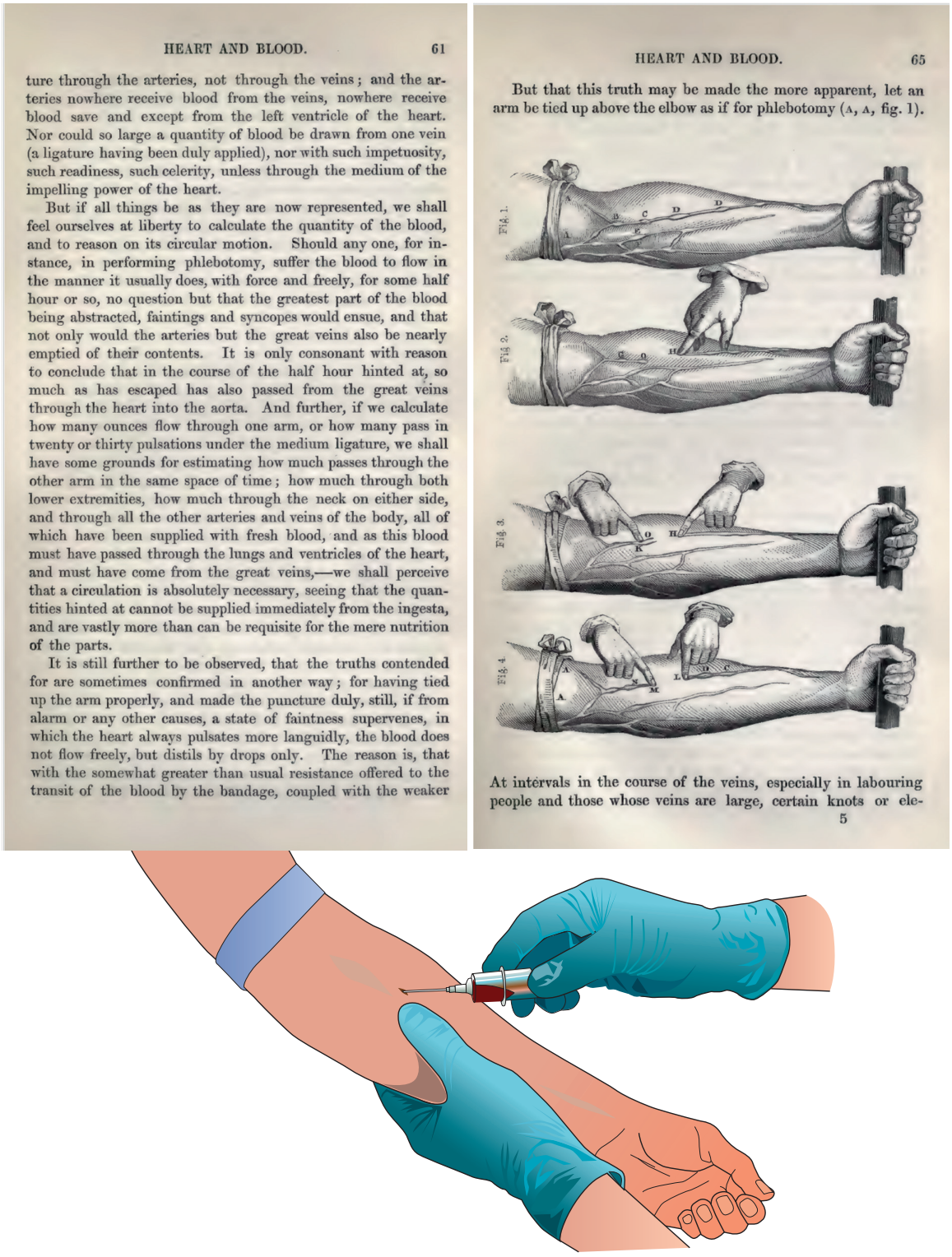
Galen, the great physician of antiquity, had taught that blood was continuously created in the liver and consumed by the body. Harvey, after anatomical dissections and careful

observations of living animals, doubted this. But doubt alone was not proof. According to Galen's model, the liver manufactured new blood from digested food as needed, while the tissues 'burned up' or consumed this blood, with no concept of recirculation. Galen also believed that blood seeped from the right side of the heart to the left through invisible pores in the septum, not that the heart actively pumped it. Blood movement to and from the heart was not seen as a continuous closed loop, but more as one-way movement and consumption. This view was deeply entrenched in both medical teaching and practice for centuries. Harvey turned to numbers. He asked: if blood is continually created anew, how much would need to be made every hour to supply the heart's pumping? And if so much blood is truly consumed, where does it all go?

Harvey's revolutionary insight emerged from a series of elegant measurements and order-of-magnitude calculations that revealed the impossibility of the prevailing Galenic model. Working with sheep, pigs, and other mammals, Harvey first measured the total volume of blood that could be drained from these animals after death. He found that a typical sheep contained roughly a few liters of blood, just shy of the 5-6 liters of blood in the typical adult human body.

In his treatise, Harvey offered a radically simple argument. He proposed that the blood could not possibly be produced fast enough to account for how much the heart ejects. The only possibility was that it recirculated. He never presented a formal equation, but he offered an invitation for the reader to do the estimate for themselves, an invitation we accept below. Here is the key passage from *Exercitatio Anatomica de Motu Cordis et Sanguinis in Animalibus* (also shown in Figure 1) that invites us to take up pencil and paper in hand and to calculate:

> But if all things be as they are now represented, we shall feel ourselves at liberty to calculate the quantity of the blood, and to reason on its circular motion. Should any one, for instance, in performing phlebotomy, suffer the blood to flow in the manner it usually does, with force and freely, for some half hour or so, no question but that the greatest part of the blood being abstracted, faintings and syncopes would ensue, and that not only would the arteries but the great veins also be nearly emptied of their contents. It is only consonant with reason to conclude that in the course of the half hour hinted at, so much as has escaped has also passed from the great veins through the heart into the aorta. And further, if we calculate how many ounces flow through one arm, or how many pass in twenty or thirty pulsations under the medium ligature, we shall have some grounds for estimating how much passes through the other arm in the same space of time ; how much through both lower extremities, how much through the neck on either side, and through all the other arteries and veins of the body, all of which have been supplied with fresh blood, and as this blood must have passed through the lungs and ventricles of the heart, and must have come from the great veins, – we shall perceive that a circulation is absolutely necessary, seeing that the quantities hinted at cannot be supplied immediately from the ingesta, and are vastly more than can be requisite for the mere nutrition of the parts.



HEART AND BLOOD. 61

ture through the arteries, not through the veins; and the arteries nowhere receive blood from the veins, nowhere receive blood save and except from the left ventricle of the heart. Nor could so large a quantity of blood be drawn from one vein (a ligature having been duly applied), nor with such impetuosity, such readiness, such celerity, unless through the medium of the impelling power of the heart.

But if all things be as they are now represented, we shall feel ourselves at liberty to calculate the quantity of the blood, and to reason on its circular motion. Should any one, for instance, in performing phlebotomy, suffer the blood to flow in the manner it usually does, with force and freely, for some half hour or so, no question but that the greatest part of the blood being abstracted, faintings and syncopes would ensue, and that not only would the arteries but the great veins also be nearly emptied of their contents. It is only consonant with reason to conclude that in the course of the half hour hinted at, so much as has escaped has also passed from the great veins through the heart into the aorta. And further, if we calculate how many ounces flow through one arm, or how many pass in twenty or thirty pulsations under the medium ligature, we shall have some grounds for estimating how much passes through the other arm in the same space of time; how much through both lower extremities, how much through the neck on either side, and through all the other arteries and veins of the body, all of which have been supplied with fresh blood, and as this blood must have passed through the lungs and ventricles of the heart, and must have come from the great veins,—we shall perceive that a circulation is absolutely necessary, seeing that the quantities hinted at cannot be supplied immediately from the ingesta, and are vastly more than can be requisite for the mere nutrition of the parts.

It is still further to be observed, that the truths contended for are sometimes confirmed in another way; for having tied up the arm properly, and made the puncture duly, still, if from alarm or any other causes, a state of faintness supervenes, in which the heart always pulsates more languidly, the blood does not flow freely, but distils by drops only. The reason is, that with the somewhat greater than usual resistance offered to the transit of the blood by the bandage, coupled with the weaker

HEART AND BLOOD. 65

But that this truth may be made the more apparent, let an arm be tied up above the elbow as if for phlebotomy (A, A, fig. 1).

Fig. 1.

Fig. 2.

Fig. 3.

Fig. 4.

At intervals in the course of the veins, especially in labouring people and those whose veins are large, certain knots or ele-

5

Figure 1: William Harvey and the circulation of blood. Left: A page from William Harvey's "An Anatomical Disquisition on the Motion of the Heart and Blood in Animals." Harvey provides the factual backdrop and the concept of an estimate for quantifying the blood traveling throughout a human body. Right: Drawing illustrating how Harvey estimated the blood flow in an arm.

Your job in this problem is to use what you know about how much blood they take from you in a typical blood test and how long it takes them to do it to make a very naive estimate for the volume of blood that must pass from the great veins through the heart into the aorta each day. Compare that number to the total blood volume in the body and to any plausible production rate from food. Follow Harvey by starting from a single arm and scaling up to the whole body by a factor you defend at the order-of-magnitude level. Treat your blood-draw-based estimate as a strict lower bound because the measurement setup throttles the flow, and note explicitly that even this lower bound is enough for Harvey's contradiction. As a modern extension, give a second, independent estimate of cardiac output using heart rate and stroke volume (or blood pressure arguments) and reconcile the two. Comment on the Galenesque and Harveyesque pictures. The main point here is not to nail the pumped volume precisely, but to see that even a naive lower bound already far exceeds what could be supplied from the ingesta if there were no circulation.

Solution: A typical blood test takes around the following amount of blood,

$$V_{\text{draw}} \approx f \times 10 \text{ mL}$$

and takes about

$$t_{\text{draw}} \approx f \text{ min.}$$

This corresponds to a lower bound for the blood flow rate in the vein of

$$Q_{\text{vein}} = \frac{V_{\text{draw}}}{t_{\text{draw}}}.$$

We scale from one arm to the whole body. In my arm, I count at least around ten veins. Then the heart is also supplying the trunk, the other arm, the legs... I can assume that as my legs are wider I might have more veins in them. Therefore, we take for the flow rate in the whole body,

$$Q_{\text{body}} \approx 100 \times Q_{\text{vein}}.$$

The volume of blood pumped per day is then

$$V_{\text{day}} \approx Q_{\text{body}} \times (1 \text{ day}) = 100 \times \frac{V_{\text{draw}}}{t_{\text{draw}}} \times (1 \text{ day}),$$

which gives

$$V_{\text{day}} \approx 100 \times \frac{f \times 10 \text{ mL}}{f \times \text{ min}} \times (24 \times 60\text{min}) \approx 10^6\text{mL} \approx 10^3 \text{ L}.$$

We compare this daily pumped volume to the total blood volume in the body,

$$V_{\text{blood}} \approx 5 \text{ L}.$$

We have that $V_{\text{day}} \gg V_{\text{blood}}$, the same blood must we be passing through the heart many times per day. Especially since we took a lower bound of the blood flow rate. Continuous

production and consumption of blood from food, would necessitate a few hundreds liters of blood to be produced every day!

Indeed, say we ingest a kilogram of food a day and a few liters of water then, taking for blood density, water density, this would give us around a few liters of blood produced a day, which is far from the 1000 liters of blood circulating through our heart every day!

We now look at cardiac output using heart rate and stroke volume. A typical heart rate is

$$r \approx 60 \text{ min}^{-1},$$

and a typical stroke volume is

$$V_{\text{stroke}} \approx 100 \text{ mL}.$$

The cardiac output is therefore

$$Q_{\text{body}}^{\text{cardiac out}} \approx r \times V_{\text{stroke}},$$

and the total volume pumped per day is

$$V_{\text{day}}^{\text{cardiac out}} \approx Q_{\text{body}}^{\text{cardiac out}} \times (1 \text{ day}) = V_{\text{stroke}} \times r \times (1 \text{ day}).$$

Which gives,

$$V_{\text{day}}^{\text{cardiac out}} = 100 \text{ mL} \times 60 \text{min}^{-1} \times (24 \times 60 \text{ min}) \approx 10^7 \text{ mL} = 10^4 \text{L}.$$

We have an extra factor 10 compared to the approximation with the blood draw. This makes sense, since the later approximation was a lower bound.

The Galenesque picture is not compatible with any plausible blood rate production from food. Blood has to be "recycled" and is pumped several times a day through the heart.

3. 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 work out the thickness of the oil film (the height of a lipid!) that covered the surface of Clapham common pond. Does your number jibe with what you know about the structure of lipids?

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:

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

(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 as shown in Figure 3?

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:

$$\text{few g} \times \frac{\text{mol}}{1000\ \text{g}} \times \frac{6 \times 10^{23}\ \text{lipid molecules}}{\text{mol}} \approx 2 \times 10^{21}\ \text{lipids}. \tag{3}$$

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:

$$\text{area} = \frac{\text{few} \times 10^3\ \text{m}^2}{2 \times 10^{21}\ \text{lipids}} \approx \frac{10^{-18}\ \text{m}^2}{\text{lipid}} \times \frac{10^{18}\ \text{nm}^2}{\text{m}^2} = 1\ \text{nm}^2/\text{lipid}. \tag{4}$$

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.

5. The concentration rule of thumb.

(a) One of the key rules of thumb we will invoke over and over again is knowledge of the concentration corresponding to one molecule per *E. coli* cell. Using that the volume of such a cell is approximately 1 μm^3, 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.

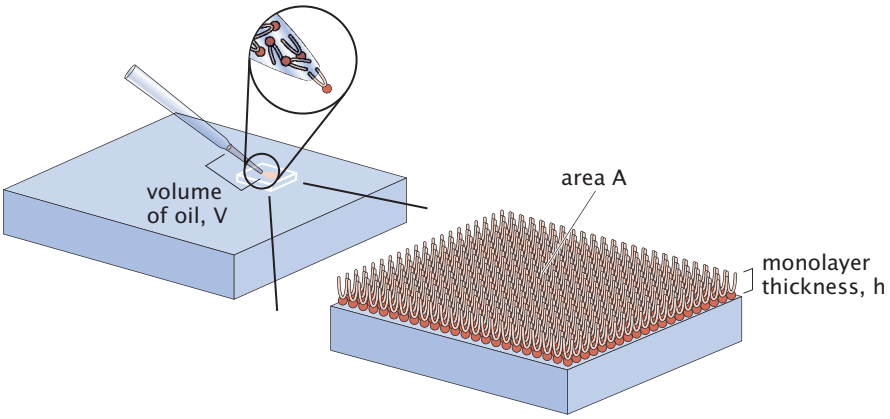


Figure 2: Putting oil on water to measure molecular dimensions. Here we see that the lipid molecules form a monolayer.

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.

Solution: Given the volume of approximately 1 $\mu\text{m}^3 = 10^{-15}\,\text{L}$ derived in part (a) of Part 4., it follows that the concentration of 1 molecule per *E. coli* cell should be

$$[1 \text{ molecule}/\textit{E. coli} \text{ cell}] \approx \frac{10^{15}\,\text{molecules}}{L} \cdot \frac{\text{mol}}{6 \times 10^{23}\,\text{molecules}} \approx 1\,\text{nM}. \tag{5}$$

(b) As an application of this idea, how many H^+ ions are there in a bacterial cell if the pH is 7.0? State any assumptions you make (for example, that the pH is uniform throughout the cell and that you can ignore buffering for the purpose of this estimate).

Solution: We convert pH to concentration of H^+ ions from the relationship

$$[H^+] = 10^{-\text{pH}}\ \text{M}. \tag{6}$$

So for a pH of 7.0, $[H^+] = 10^{-7}\,\text{M}$, or $10^2\,\text{nM}$. Thus, since we know that 1 molecule per bacterium is equivalent to a concentration of ≈ 1 nM, a pH of 7.0 corresponds with 100 H^+ ions in the cell.

(c) It is very useful to have a sense of how far molecules are apart at a given concentration. Work out a formula that relates the typical spacing between molecules d to the concentration c by assuming the molecules are uniformly distributed in 3D. Then, make a plot that shows d as a function of c for concentrations ranging from nM to M. Make sure your axes are labeled with units.

Solution: Our goal is to determine the average spacing between molecules given a specified concentration. From the molar concentration, we can express the molecular density as

$$\frac{c\,\text{mol}}{L} \cdot \frac{6 \times 10^{23}\,\text{molecules}}{\text{mol}} \cdot \frac{\text{L}}{10^{24}\,\text{nm}^3} = c\ 0.6\,\text{molecules}/\text{nm}^3. \tag{7}$$

Inverting this result, we generate the volume of solution occupied by each molecule at molar concentration C M,

$$V = 1.66\,\text{nm}^3/\text{molecule}\ 1/c. \tag{8}$$

The cubed root of this volume thus indicates the average separation between molecules, such that we may conclude

$$d \propto V^{1/3} = \frac{1.18}{c^{1/3}}\,\text{nm}. \tag{9}$$

(d) As an application of your thinking from part (c), explain what the concept of the "critical concentration" is for the polymerization of actin filaments. Then, provide a rough estimate of the mean spacing between actin monomers in a solution at the critical concentration. State and justify the critical concentration you use (from memory or by citing a source).

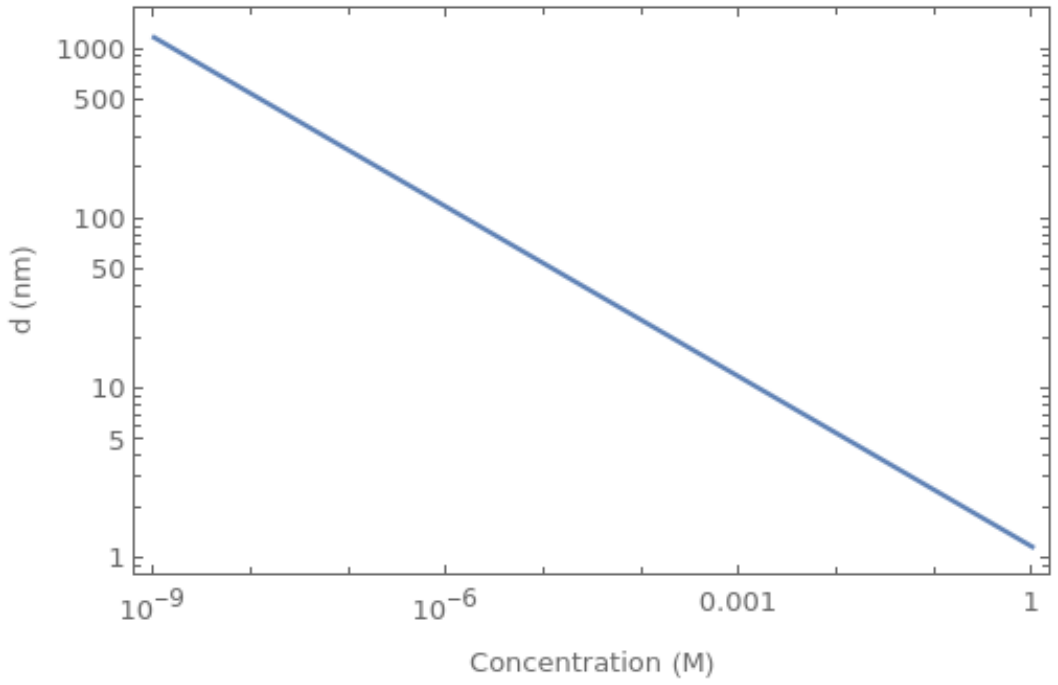


Figure 4: Plot of average separation d between molecules against concentration ranging from nM to M.

Solution: Polymerization in solution can be considered under a simple model: after a nucleation phase where 3-4 monomers randomly interact to form a nucleus which can then be elongated. This elongation is then governed by monomer capture events (monomers polymerize onto one end of the polymer) and monomer escape events (a monomer leaves one end of the polymer). Monomer capture is dependent on the interaction between one end of the polymer and a monomer in solution, and is therefore likelier the smaller the mean separation between monomers in solution - we can capture this by setting the rate of monomer capture to be proportional to the concentration of monomers. Monomer escape, however, does not require interaction with monomers in solution and is therefore independent of the monomer concentration. This can be written as

$$\frac{dn}{dt} = k_{on}C - k_{off} \tag{10}$$

where n is the number of monomers that constitute the polymer we are considering, k_{on} (resp. k_{off}) are rate constants for monomer capture (resp. escape), and C is the monomer concentration in solution.

We can then see that the system is in steady state when $C = C_{crit} \equiv \frac{k_{off}}{k_{on}}$. This is referred to as the critical concentration - below this concentration the filaments depolymerize and above this concentration the filaments keep polymerizing, pulling monomers out of solution until the critical concentration is reached.

Actin polymerization, however, is more complicated than this draft model captures. It turns out that the ends of actin filaments behave asymetrically - i.e. actin filaments have a 'plus'

end and a 'minus' end with the 'plus' end having a higher growth (and shrinkage) rate. A better first-order model is given by

$$\frac{dn}{dt} = k_{on}^{+}C + k_{on}^{-}C - k_{off}^{+} - k_{off}^{-} \tag{11}$$

where the '+' and '-' superscripts in the rate constants denote the 'plus' and 'minus' end of the filament respectively. (Note: There are more subtleties involved in actin polymerization - e.g. the monomer capture rates also depend on whether the monomer is ADP or ATP bound - but the above is sufficient as a first-order model.)

This system now has three critical concentrations: $C_+ \equiv \frac{k_{off}^{+}}{k_{on}^{+}}$ below which both ends shrink, $C_- \equiv \frac{k_{off}^{-}}{k_{on}^{-}}$ above which both ends grow, and $C_{TM} \equiv \frac{k_{off}^{+}+k_{off}^{-}}{k_{on}^{+}+k_{on}^{-}}$ at which the system reaches steady-state (referred to as 'treadmilling') whereby the 'plus' end grows at the same rate at which the 'minus' end shrinks. For concentrations between C_- and C_+, the 'plus' end grows and the 'minus' end shrinks, with the relative rates of the two processes determining whether or not the filament elongates.

As per BNID 112788, $C_+ \sim 0.06$ μM and $C_- \sim 0.6$ μM, corresponding to a mean separation of ~ 300 nm and ~ 100 nm respectively. C_{TM} lies somewhere between the two, which we estimate to be ~ 0.2 μM and which corresponds to a mean separation of ~ 200 nm.

6. 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 roughly 72,000 ribosomes per cell, and slow growing *E. coli* with a division time of 100 minutes have about an order of magnitude 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 5, 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. State clearly what region you counted, what depth (or slice thickness) you implicitly assumed, and how you handled the green and yellow ribosomes.

Solution: In the close up view of the 3D reconstruction (panel C of the Fig 5) 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.

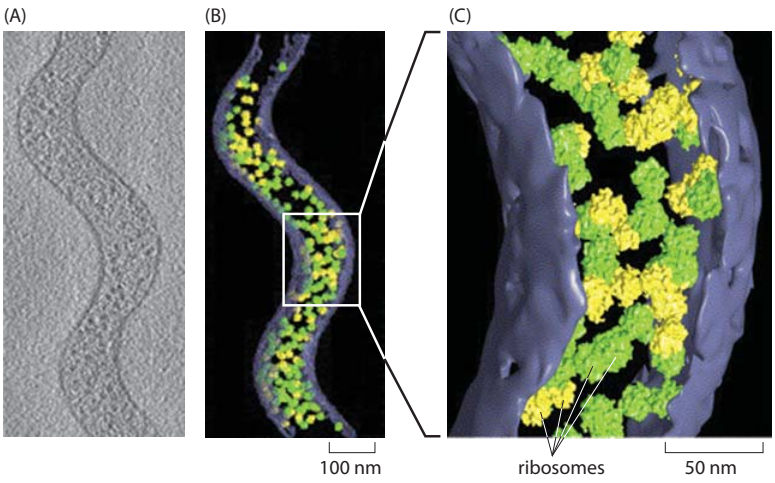


Figure 5: Cryo EM study of a bacterial cell. These images are of the tiny bacterium, *Spiroplasma melliferum*. Using algorithms for pattern recognition and classification, components of the cell such as ribosomes were localized and counted. (A) Single cryo-electron microscopy image. (B) 3D reconstruction 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).

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

$$\begin{aligned} V_{\text{close up}} &\approx \frac{\pi \times (100\,\text{nm})^2}{4} \times 200\,\text{nm} \\ &\approx 2 \times 10^6\,\text{nm}^3. \end{aligned} \tag{12}$$

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

$$\begin{aligned} \rho &= \frac{N_{\text{close up}}}{V_{\text{close up}}} \\ &= \frac{35}{2 \times 10^6\,\text{nm}^3} \\ &\approx 2 \times 10^{-5}\,\text{nm}^{-3} \\ &= 2 \times 10^{-5}\,\text{nm}^{-3} \times \left(\frac{10^3\,\text{nm}}{1\,\mu\text{m}}\right)^3 \\ &= 2 \times 10^4\,\mu\text{m}^{-3}. \end{aligned} \tag{13}$$

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) Now that you have figured out the number of ribosomes per μm^3, use that result to estimate how long it would take to reproduce the entire proteome of a bacterial cell. To that end, you will need to supply a typical volume of a bacterial cell, its mass, what fraction of that mass is protein, and then use the known rate of translation (see Bionumbers) to estimate how long it takes to double the proteome given the number of ribosomes. State and justify any assumptions you make about typical protein size (mass or length) and about whether all ribosomes are actively translating. Make sure you explain your arguments and the insights they provide.

Solution: From part (a) we have a ribosome density ρ. For a cell of volume V_{cell}, the number of ribosomes is

$$N_{\text{ribo}} \approx \rho\, V_{\text{cell}}. \tag{14}$$

Assume the cell has the same density as water, so its total mass is

$$m_{\text{cell}} \approx \rho_{\text{w}}\, V_{\text{cell}}. \tag{15}$$

We take the dry mass to be one third of the total mass, and we assume that the dry mass is only proteins or nucleic acids. We consider that half of the dry mass is protein, such as

$$m_{\text{protein}} \approx \frac{1}{2}\, m_{\text{dry}} \approx \frac{1}{2}\left(\frac{1}{3} m_{\text{cell}}\right) = \frac{1}{6}\rho_{\text{w}} V_{\text{cell}}. \tag{16}$$

We computed before the typical molecular mass of an amino acid $M_{\text{aa}} \approx 100\,\text{Da} \approx 100\,\text{g/mol}$. The number of amino acids in the whole proteome is therefore,

$$N_{\text{aa}} \approx \mathcal{N}_A \frac{m_{\text{protein}}}{M_{\text{aa}}} = \mathcal{N}_A \frac{\rho_{\text{w}} V_{\text{cell}}}{6\, m_{\text{aa}}}, \tag{17}$$

with $\mathcal{N}_A \approx 6 \times 10^{23}$ mol^{-1}, the Avogadro number. To double the proteome, the cell must synthesize roughly this many amino acids again. If each active ribosome elongates at rate v (aa/s), and only a fraction ϕ of ribosomes are actively translating, then the amino-acid production rate is

$$r_{\text{aa}} \approx \phi N_{\text{ribo}}\, v \approx \phi \rho V_{\text{cell}}\, v. \tag{18}$$

Therefore the time to reproduce the entire proteome is

$$t_{\text{proteome}} \approx \frac{N_{\text{aa}}}{r_{\text{aa}}} \approx \frac{\rho_{\text{w}} V_{\text{cell}}/(6 m_{\text{aa}})}{\phi \rho V_{\text{cell}} v} = \frac{\mathcal{N}_A \rho_{\text{w}}}{6\, M_{\text{aa}}\, \phi\, \rho\, v}. \tag{19}$$

We use the following parameter values,

$\rho \approx 2 \times 10^4\ \mu\text{m}^{-3}$,
$v \approx 10$ aa/s, Peptite chain elongation in *E. coli*: 12-21 aa/s (BNID 100059),
$\phi \approx 0.8$,
$\rho_{\text{w}} \approx 1\ \text{g} \cdot \text{cm}^{-3} = 10^{-12}\ \text{g} \cdot \mu\text{m}^{-3}$,
$M_{\text{aa}} \approx 10^2$ Da.

The final result is,

$$t_{\text{proteome}} \approx \frac{10^{-12}\,\text{g}\cdot\mu\text{m}^{-3} \times 6 \times 10^{23}\,\text{mol}^{-1}}{6 \times 10^{2}\,\text{g mol}^{-1} \times 0.8 \times 2 \times 10^{4}\,\mu\text{m}^{-3} \times 10\,\text{s}^{-1}} \tag{20}$$

$$\approx 6 \times 10^{3}\,\text{s} \approx 2\,\text{hours}. \tag{21}$$

7. Sizing up the Central Valley.

In this course, we are going to consider biological phenomena across a huge range of scales in space and time, including examining the ways in which the biology of our planet is altered by humans. 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 6. 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 × 20 cm and we would need a cup of water (250 mL) everyday to irrigate the flower. That would give us

$$\text{water needed to irrigate crops in a unit area per day} = \frac{0.25\ \text{L/day}}{20\ \text{cm} \cdot 20\ \text{cm}} = \frac{0.25\ \text{L/day}}{0.04\ \text{m}^2} \approx 5\ \text{L/m}^2 \cdot \text{day}. \tag{22}$$

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

$$\underbrace{10^{10}\ \text{m}^2}_{\text{Central Valley area}} \times 5\,\text{L/m}^2 \cdot \text{day} \times 30\ \text{days/month} \times \underbrace{9\ \text{months/year}}_{\text{crop season}} \approx 10^{13}\text{L/year}. \tag{23}$$

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.

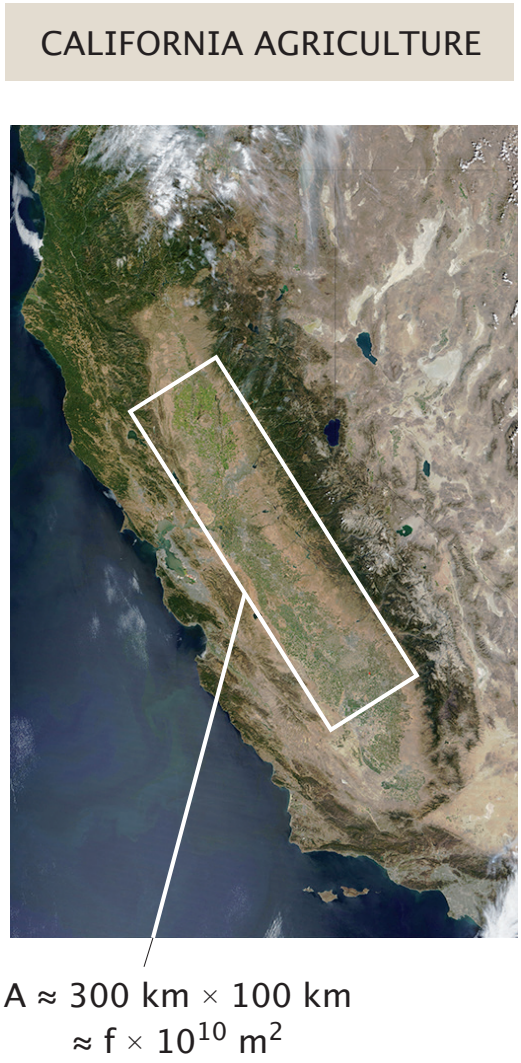


Figure 6: Satellite image of California's Central Valley.

(b) Nitrogen usage. Since the beginning of the twentieth century, synthetic nitrogen fixation via the Haber–Bosch process has enabled the modern world to feed of order half of humanity. 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. State clearly what you mean by "fertilizer used" (for example, kg of nitrogen applied per year).

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

$$\text{Biomass per square metre} \approx 10 \text{ kg/m}^2. \tag{24}$$

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

$$\text{Nitrogen per biomass per square metre} = 0.5\% \times 10 \text{ kg/m}^2 = 0.05 \text{ kg/m}^2 \tag{25}$$

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

$$0.05 \text{ kg fertilizer/m}^2 \times 10^{10} \text{ m}^2 = 5 \times 10^8 \text{ kg fertilizer}. \tag{26}$$

(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. State clearly what you are counting (for example, total mass of active ingredient applied per 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 $m^3 = 1000$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:

$$\text{Pesticide needed every year} = \frac{1000\text{L/year}}{1\text{km} \cdot 1\text{km}} = 1 \times 10^{-3}\ \text{L/m}^2 \cdot \text{year}. \tag{27}$$

For the entirety of the Central Valley, we need

$$10^{10}\ \text{m}^2 \times 10^{-3}\ \text{L/m}^2 \cdot \text{year} = 10^7\ \text{L/year}. \tag{28}$$

Assuming that the density of pesticide is the same as water ($\rho = 1$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. Cite your sources clearly.

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} $m^3 = 3 \times 10^{13}$ 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.