BE/APh161: Physical Biology of the Cell
Homework 8
Due Date: Wednesday, March 18, 2020

"How can the events in *space and time* which take place within the spatial boundary of a living organism be accounted for by physics and chemistry?" - Erwin Schrödinger **What is Life?**

This final homework is a celebration of physical biology, drawing its inspiration from disparate puzzles about the living world while trying to showcase some of the many key principles that have come up in the course. What are those key principles? 1) Wilson, order of magnitude thinking, street fighting-mathematics and figuring out estimates about the world around us without looking anything up. I argue that this is a skill that transcends this course. Rather, it is one of the key tools that we should all carry around as our route to answering the question: what sets the scale of X? 2) The great probability distributions as a window onto mechanism. We have seen over and over that probability serves as the natural language of biology, whether in thinking about the waiting time for the next step of a molecular motor or the distribution of mRNA molecule counts in a population or the frequency of alleles. 3) Physical biology superpowers - the key protocols of dynamics, continuum theory and statistical mechanics. One of our most important visions was that of field theory, one of the greatest inventions in the history of science.

1. Physical Biology of Viruses.

Since their discovery over 100 years ago, viruses have always occupied a central place in biology. One of the debates that has swirled around their existence is the simple question of whether or not they are "alive". During the Max Delbrück era, he was interested in finding the "hydrogen atom" of life and found bacterial viruses (the so-called bacteriophages - literally, bacteria eaters) would serve perfectly in that capacity. Several years ago when I was teaching this course, the big news was Zika virus and before that a very scary Ebola outbreak. This year, we are faced with increasingly alarming news of the coronavirus, COVID-19. My own switch from condensed matter physics to biology was partly elicited by an amazing paper from the group of

virus	size (nm)	genome size (base pairs)	genome type, capsid structure	BNID
porcine circovirus (PCV)	17	1,760	circular ssDNA, icosahedral	106467, 106468
cowpea mosaic virus (CPMV)	28	9,400	2 ssRNA molecules, icosahedral	106454, 106455
cowpea chlorotic mottle virus (CCMV)	28	7,900	3 ssRNA molecules, icosahedral	106456, 106457
φX174 (*E. coli* bacteriophage)	32	5,400	ssDNA, icosahedral	103246, 106442
tobacco mosaic virus (TMV)	40×300	6,400	ssRNA, rod shaped	104376, 104375, 106453
polio virus	30	7,500	ssRNA, icosahedral	103114, 111324
φ29 (*Bacillus* phage)	45x54	19,000	dsDNA, icosahedral (T3)	109734
lambda phage	58	49,000	dsDNA, icosahedral (with tail)	103122, 105770
T7 bacteriophage	58	40,000	dsDNA, 55 genes, icosahedral (T7)	109732, 109733
adenovirus (linear DNA)	88-110	36,000	dsDNA, icosahedral	103114, 103115, 106441
influenza A	80-120	14,000	ssRNA, roughly spherical	104073, 105768
HIV-1	120-150	9,700	ssRNA, roughly spherical	101849, 105769
herpes simplex virus 1	125	153,000	dsDNA, icosahedral	103114, 106458
Epstein-Barr virus (EBV)	140	170,000	dsDNA, icosahedral	103246, 111424
mimivirus	500	1,200,000	dsDNA, icosahedral	105142, 105143
pandora virus	500x1000	2,800,000	dsDNA, icosahedral	109554, 109556

Figure 1: Sizes of viruses. The table considers both RNA and DNA viruses and reports both the size of the virion and the length of the genome.

Carlos Bustamante at UC Berkeley who had figured out a way to measure the build up of pressure as DNA is packed into the bacteriophage capsid. In this problem, we take a random walk through the physical biology of viruses, honoring them as one of the most sophisticated, interesting and scary parts of the biological world.

(A) Let's begin by considering the data storage capacity of viruses. Choose an RNA virus (such as influenza, HIV or COVID-19) and a bacteriophage (such as lambda or T4) and compute the physical data storage capacity when their genomes are packed within the virion. Figure 1 is a resource that will allow you to understand viral sizes. I am talking about units of bits/μm^3. Compare this to the 4 TB hard drive that one uses to back up a laptop. How many viruses would it take to store the entirety of the Library of Congress? How much volume would all of those viruses take up?

(B) One of the most important properties of a given infection is the so-called "burst size", the number of new viruses produced per infected cell. One of the

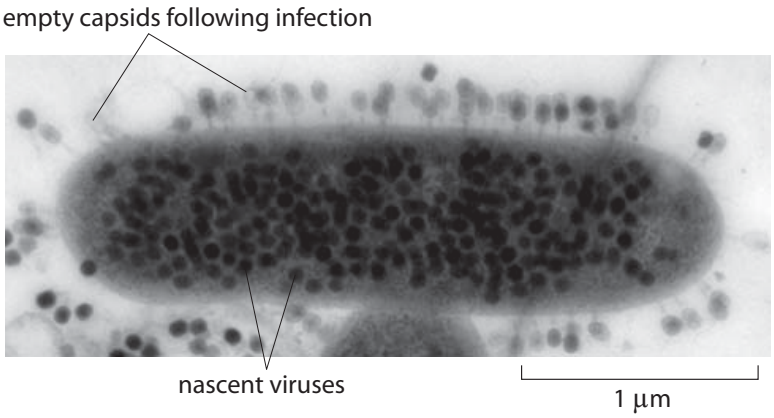
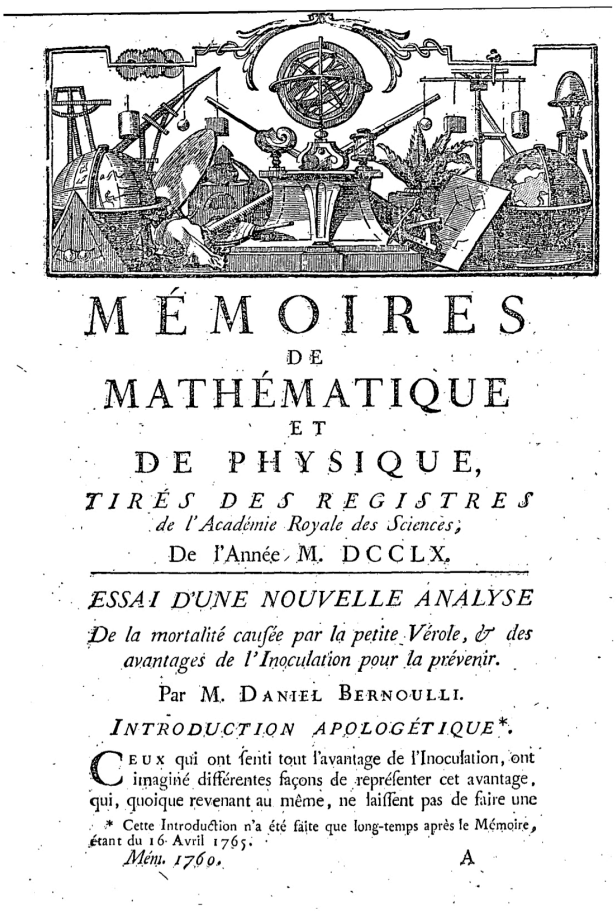


Figure 2: Synthesis of new viruses in an infected bacterium.

original hypotheses (which you will refute here) for what controls the burst size is the available volume within the host cell. Given that for a typical bacteriophage infection the burst size is roughly 100 viruses, what fraction of the volume is taken up by the newly synthesized viruses? Figure 2 shows an electron microscopy image of an infected bacterium.

(C) A simple model of viral spread through a population is the so-called SIR model, where S refers to susceptible, I refers to infected and R refers to recovered-removed. There is a long tradition running all the way back to the Greeks of trying to understand the population dynamics of disease spread. In 1760, the great Daniel Bernoulli mused on the topic of small pox and vaccinations as shown in Figure 3. If the rate of infection is given by r, and the rate of recovery is given by α, write three coupled differential equations for the dynamics of S, I and R. Choose reasonable values of those parameters based on looking at the current story of Coronavirus. Consider a closed and isolated city such as Wuhan, China (that is, make the clearly overly optimistic assumption that no one leaves or enters the city) and solve for the three variables as a function of time and plot them together on a common plot such as that shown in Figure 5. Explain the phase portrait in detail that is shown in Figure 4. Specifically, find an analytic expression for the parameter ρ which is the critical population size such that $dI/dt > 0$.

(D) How are viruses transmitted? Three key routes are through the respira-



MÉMOIRES
DE
MATHÉMATIQUE
ET
DE PHYSIQUE,
TIRÉS DES REGISTRES
de l'Académie Royale des Sciences;
De l'Année M. DCCLX.

ESSAI D'UNE NOUVELLE ANALYSE
De la mortalité causée par la petite Vérole, & des avantages de l'Inoculation pour la prévenir.

Par M. DANIEL BERNOULLI.

INTRODUCTION APOLOGÉTIQUE.*

CEUX qui ont senti tout l'avantage de l'Inoculation, ont imaginé différentes façons de représenter cet avantage, qui, quoique revenant au même, ne laissent pas de faire une

* Cette Introduction n'a été faite que long-temps après le Mémoire, étant du 16 Avril 1765.

Figure 3: Paper from Daniel Bernoulli, 1760 in which he considered a dynamic model of small pox.

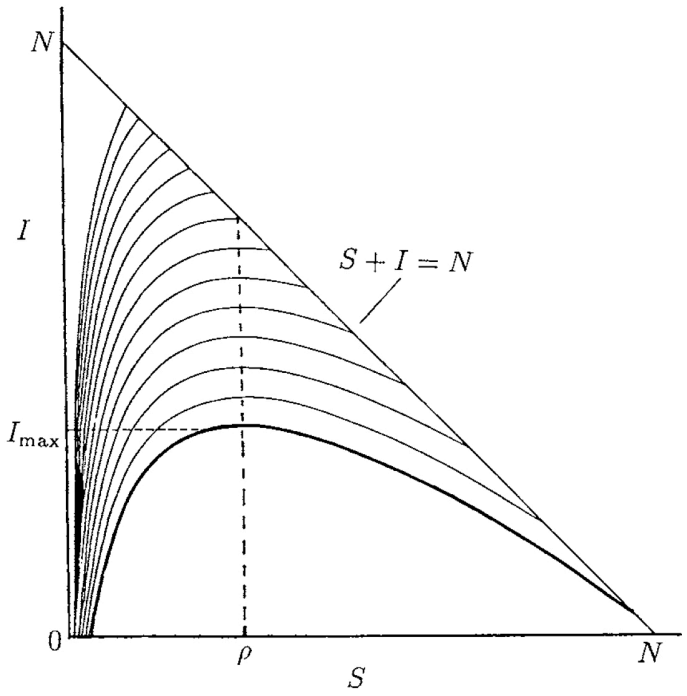


Figure 4: Phase portrait in the SIR model. The total population size is N. The number of susceptible individuals plus the number of infected individuals at $t = 0$ is equal to the population size. Thus, all subsequent evolution of the population must exist within the triangle since $S(t) + I(t) + R(t) = N$. ρ is the critical population size for the occurrence of an epidemic.

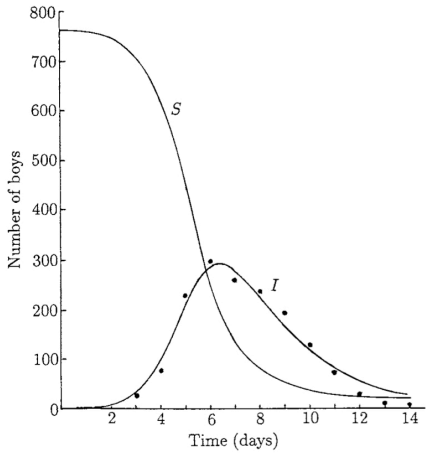


Figure 5: Influenza data from a boarding school for boys reported in *The Lancet* in 1978.

tory tract, the digestive tract and the reproductive tract. In all three cases, our bodies are set up with a number of different tricks to resist infection including mucus and ciliary transport in our respiratory and digestive tracts and harsh conditions in our digestive tract such as low pH. The current coronavirus epidemic is apparently passed through the respiratory tract and in this part of the problem, we appeal to Figure 6 for a look at the distribution of droplet sizes. The claim is that a strong sneeze or cough can contain more than 10,000 such droplets. How much volume is that? Does that make sense? Estimate how many influenza or coronavirus particles will be carried in a typical droplet. I have not done all of these estimates carefully enough for my own satisfaction so this part of the problem is an adventure for all of us. A very interesting source of information on this is the work of Prof. Lydia Bourouiba from MIT who does visualization experiments on humans coughing.

2. Waiting time distributions.

One of the big messages of the course is the deep insights that come from a probabilistic assessment of biological systems. Our slogan might be: mechanistic information is hidden in the probability distributions. The binding

carrying droplet-nuclei which remained airborne after sneezing was found to decrease geometrically with time; only 4 % remained airborne after 30 min. and 2 % after 40 min.

In the present investigation, the droplet-nuclei produced in speaking, in coughing and in sneezing have been measured by a new technique, namely, by direct micrometry after their recovery from the air on to oiled slides. The sizes of the smaller respiratory droplets have been calculated from the sizes of these droplet-nuclei. The sizes of the larger respiratory droplets have been estimated from measurements made of stain-marks found on slides exposed directly to mouth-spray. By appropriate combination of these two sets of findings, the formulation of a comprehensive size distribution for the respiratory droplets has been attempted. The duration of aerial infection by droplet-nuclei has been observed by examination of the air at intervals after droplet-spray production, for the presence both of bacteria-carrying droplet-nuclei and of all microscopically visible droplet-nuclei.

THE MEASUREMENT OF DROPLETS AND DROPLET-NUCLEI

The following expiratory activities were tested: (1) *sneezes*, induced by snuff or by tickling the nasal mucosa with a throat swab; (2) *coughs with the mouth initially closed*, voluntarily performed with the lips, or with the tongue and the upper teeth, approximated at the start of expiration; (3) *coughs with the mouth open*, voluntarily performed with the mouth kept well open and the tongue depressed; (4) *speaking loudly one hundred words*, by counting from 'one' to 'a hundred'.

A. *The measurement of stain-marks on slides exposed directly to mouth-spray*

In order that even the smallest droplet-marks might be readily visible, some dye was introduced into the mouth just prior to each test. A little congo red, eosin or fluorescein powder was applied with a throat swab to the surfaces of the mouth and fauces; the heaviest application was made to the tip of the tongue, to the front teeth and to the lips, for droplet-spray originates largely from the secretions of the anterior mouth. Following solution of the dye, droplet-spray was produced by sneezing, by coughing or by speaking; it was directed at a celluloid-surfaced slide held 3 in. in front of the mouth in tests of speaking, and 6 in. in front of the mouth in tests of coughing and sneezing. The slide was examined under the microscope, and the diameters of the first few hundred droplet-marks encountered were measured with aid of a micrometer eyepiece. In the case of each type of expiratory activity, a number of tests, from 10 to 22, were carried out, involving the measurement of 3000 droplets.

In order to ascertain the relationship between the diameters of the droplets while in their original spherical state, and the diameters of the stain-marks which the droplets leave on evaporation after impinging and flattening upon a slide, the experiments of Strausz (1926) were repeated. With the low power of a microscope and a micrometer eyepiece, large drops of saliva (1–3 mm. in diameter) were measured, first while they hung from fine glass capillaries and then again after they had fallen, flattened and evaporated on a slide. When a glass slide was used, it was found, as it had been by Strausz, that the diameters of the original droplets were about one-third those of the stain-marks. When a celluloid-surfaced slide was used, the diameters of the original droplets were about half those of the stain-marks. Celluloid slides were used throughout the present investigation, so the original droplet diameters have been calculated as half the measured diameters of the stain-marks. The size distribution so found for the droplets expelled in the different expiratory activities, is shown in Table 1. It will be noted that few droplets were found of less than 10 μ in diameter and none of less than 5 μ. It is presumed that droplets smaller than this possessed such a small mass, or evaporated rapidly to such a small mass, that they were carried past the slide in the deflected air stream.

Table 1. *The size distribution of the larger droplets*

Showing for each type of expiratory activity the diameters of 3000 droplets calculated as half the measured diameters of the stain-marks found on celluloid slides exposed a few inches in front of the mouth.

Diameter in μ	Sneezes	Coughs with mouth 'closed'	Coughs with mouth open	Speaking loudly
0–5	0	0	0	0
5–10	36	24	8	20
10–15	94	119	39	84
15–20	267	337	127	200
20–25	312	346	189	224
25–50	807	767	577	597
50–75	593	468	593	531
75–100	260	285	341	352
100–125	144	160	231	260
125–150	105	125	202	214
150–200	115	115	253	179
200–250	82	96	165	99
250–500	118	113	213	197
500–1000	59	40	52	41
1000–2000	8	5	10	2

Figure 6: Distribution of droplet sizes after a sneeze.

problems that we worked out for ligands and receptors can be thought of as giving rise to a time series that looks like a so-called telegraph signal, going back and forth between 0 and 1. Because the time of switching between bound and unbound is very fast compared to the time spent in those two states, the occupancy of the receptor is either 0 or 1.

(A) In light of this, it is interesting to explore the distribution of waiting times that we spend in the unoccupied or occupied state. To that end, we can use the interpretation of rates as follows. Consider that the receptor is currently occupied and we start a stopwatch to measure how long until a ligand hops off of it. In each instant Δt, as shown in Figure 7, there is a probability $p_+ = k_{off}\Delta t$ of hopping off of the receptor. The goal of our calculation is to work out the probability that the ligand will fall off after a time $T = n\Delta t$, where n is the number of time steps we have to wait until the ligand falls off. To do so, we imitate the figure by noting that to fall off at time T this means that the ligand will have to have *not* fallen off during all the previous steps. Since we have discretized time into slices of length Δt, show how to write the probability as a product of n independent probabilities. Use the insight that

$$lim_{n\to\infty}(1 - x/n)^n = e^{-x} \tag{1}$$

to show that the probability that the ligand falls off between time T and $T + \Delta t$ is given by

$$p(T)\Delta t = k_{off}e^{-k_{off}T}\Delta t. \tag{2}$$

Show that this probability distribution is properly normalized and then compute the average waiting time

$$\langle t \rangle = \int_0^\infty tp(t)dt. \tag{3}$$

(b) When we think about molecular motors, we will be interested in molecules that transition between more than two states, but have exponential waiting times in each of those states. Consider the case of a molecular motor that has two steps, each with a waiting time distribution that is exponential like you worked out in the first part of the problem. Using that, work out an expression for the waiting time distribution for the *composite* process made up of those two steps. That is, once again find $p(T)$ given that both t_1 and t_2 are exponentially distributed, where t_1 is the waiting time for the first step and t_2 is the waiting time for the second step. The key point in formulating

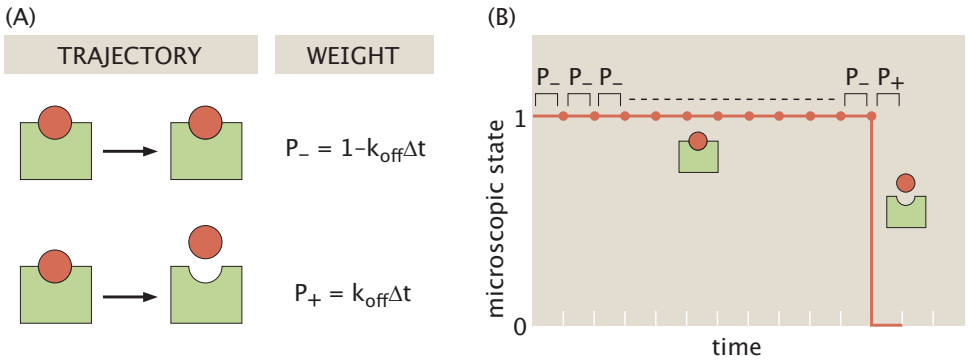


Figure 7: Computing the waiting time distribution. (A) The possible microscopic trajectories that can occur during a time step Δt. (B) Schematic of the states during all the time steps leading up to the ligand falling off of the receptor.

your thinking is that you must respect the constraint that $t_1 + t_2 = T$. Make a plot of this kind of distribution and comment on what it means.

3. MWC Ion Channel: One Equation that Rules Them All

In class, we introduced the idea of allosteric proteins as those that have a regulatory binding site that cause the protein to switch between inactive and active states. In this problem, we will take the same ideas developed in class and apply them to the so-called ligand-gated ion channels. These channels are relevant in contexts ranging from our neuromuscular junctions to the photoreceptors in our eyes to olfactory neurons. Figure 8 shows two classic examples of these channels.

(a) Write a paragraph that summarizes the function of the two ion channels shown in Figure 8. The point here is just to make sure you have a little understanding of their physiological function before we start working out their statistical mechanical properties.

(b) Make a diagram with your version of the statistical mechanics protocol showing the states and weights for the nAChR ion channel. Make sure you explain all of your notation for the parameters that appear here.

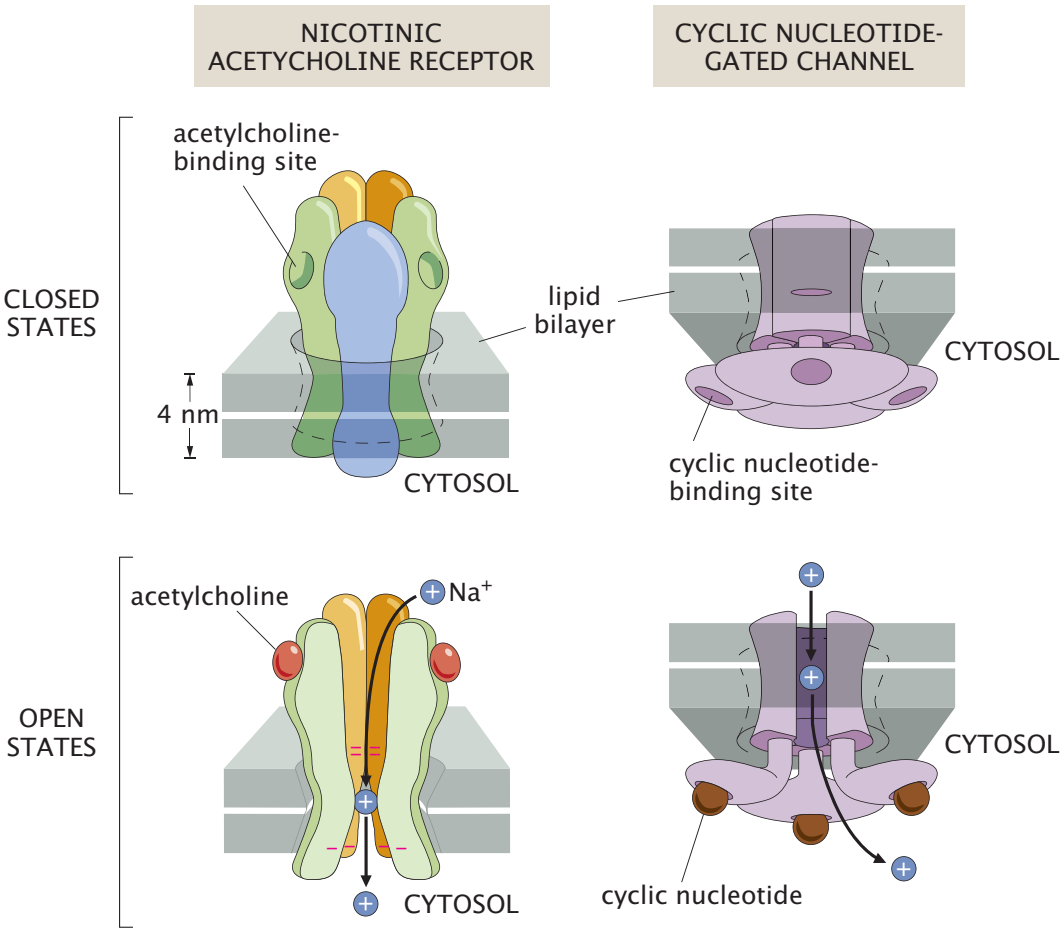


Figure 8: Key examples of ligand-gated ion channels. (left) Nicotinic acetylcholine receptor, revealing its heteropentameric structure with two binding sites for acetylcholine. (right) cGMP-gated ion channel. These channels have four cGMP binding sites.

(c) Write an equation for the probability that the channel is open $p_{open}(c)$, where c is the concentration of acetylcholine.

(c) Work out the leakiness, dynamic range and the EC50. Leakiness refers to the probability that the channel is open in the absence of ligand and can be thought of as p_{min}, the minimum probability the channel is open. Dynamic range refers to the difference between p_{max} and p_{min}, where p_{max} is the probability of being open at saturating concentrations of ligand. Find explicit expressions for both p_{min} and p_{max} and then use their difference to obtain the dynamic range. EC50 is the concentration of ligand at which the channel is halfway between p_{min} and p_{max}. Write expressions for each of the four properties listed above. Then, simplify your expressions for these various properties in the limit where $K_I/K_A >> 1$.

(d) Figure 9 shows data for the wild-type nAChR ion channel from the laboratory of our own Prof. Henry Lester. With your TA, use Digitizeit to extract the data and then make a fit using the MWC model you worked out earlier in the problem. This is Figure 1B of the paper by Labarca *et al.* included with the homework. Note that unfortunately, they chose to plot "normalized current" rather than $p_{open}(c)$. As a result, your fit will have to be to the normalized current given as

$$\text{normalized current} = \frac{p_{open}(c) - p_{min}}{p_{max} - p_{min}}. \tag{4}$$

I am excited for you to learn how to use Digitizeit because it is liberating: with it, you can take figures from anyone's papers and grab their experimental data and export it into a spreadsheet so that you can unleash your theoretical analysis on it.

4. Setting up the fly body plan.

One of the most important ideas for how positional information arises in multicellular organisms is the idea of a morphogen gradient (another serious contender is a Turing pattern). In this problem we will use a steady-state solution to the reaction-diffusion equation for Bicoid to understand how the

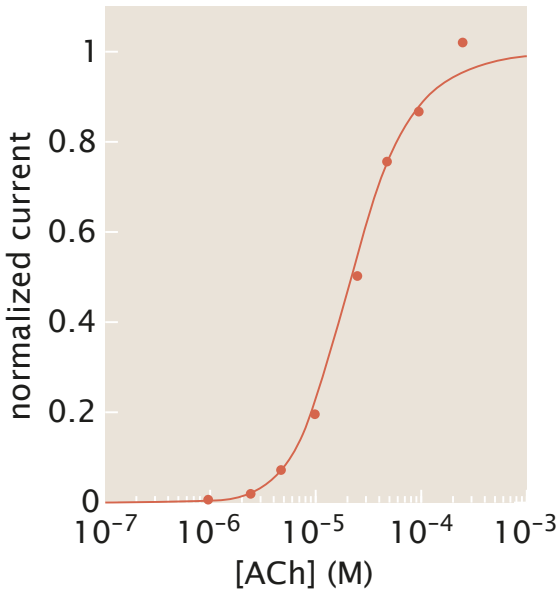


Figure 9: Ion channel currents as a function of ligand concentration. (Adapted from Labarca *et al.*, Nature, 1995).

exponential profile shown in Figure 10 is set up. Stated simply, the development of the Bicoid gradient can be thought of as resulting from a competition between the diffusion of Bicoid protein that is synthesized at the anterior end of the embryo (the mother deposits localized *bcd* mRNA there as shown in Figure 11) and the degradation of this protein while it is diffusing around.

(A) Give a brief description (a paragraph or less) of the Bicoid gradient in *Drosophila* and how it is relevant to fly development. Further, to get a feeling for the Bicoid gradient, redraw the Bicoid profile shown in Figure 10 in terms of the absolute number of Bicoid proteins per nucleus. You can make the drawing by hand or plot some approximate curve using Python. Note that there are two profiles, one for dorsal and one for ventral values of Bicoid. For the purposes of your plot, you can approximate this data as one averaged curve. To make this estimate, you will need to use the information about nuclear sizes in nuclear cycle 14 provided in Figure 4C of Gregor2007a (provided on the course website).

(B) Make a derivation of the reaction-diffusion equation and use it to justify

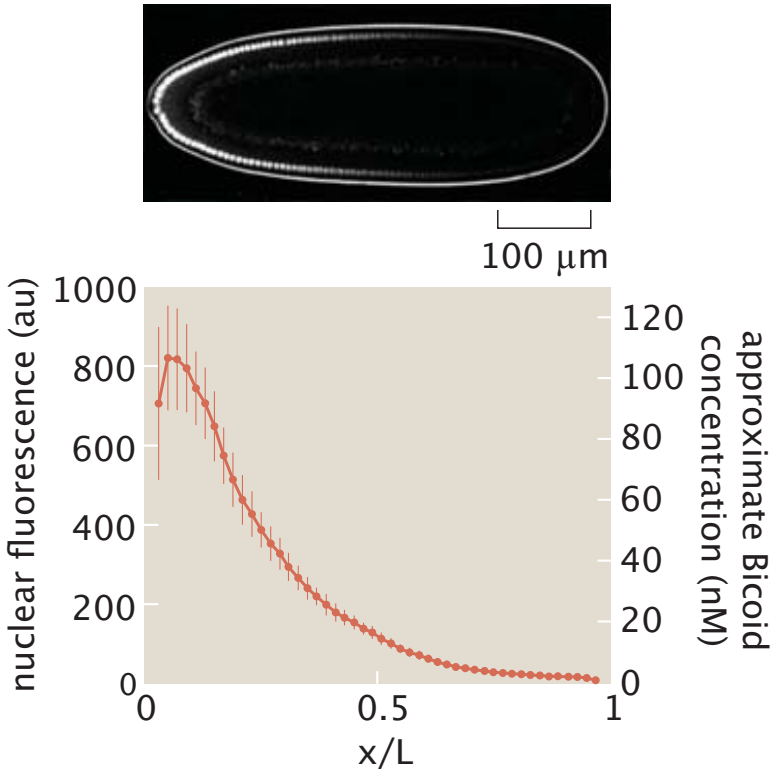


Figure 10: The Bicoid morphogen. The Bicoid activator is distributed in an exponential gradient. (Adapted from F. Liu *et al.*, Proc Natl Acad Sci USA 110:6724 2013.)

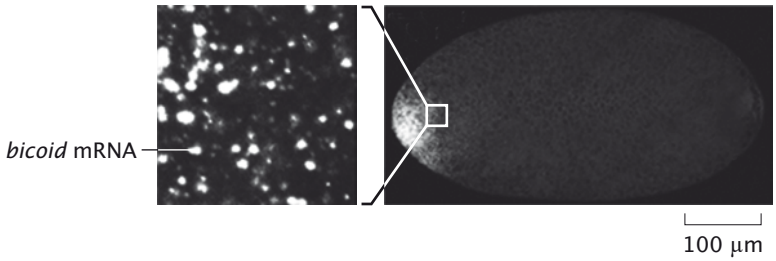


Figure 11: *bicoid* mRNA distribution. Using single molecule mRNA FISH, the localization of individual *bicoid* mRNA molecules at the anterior end of the embryo can be revealed. (Adapted from Petkova et al. (2014), *Current Biology* 24:1283.)

the form

$$\frac{\partial Bcd(x,t)}{\partial t} = D\frac{\partial^2 Bcd(x,t)}{\partial x^2} - \frac{Bcd(x,t)}{\tau}. \quad (5)$$

Make sure you explain carefully where all of these terms come from. To do so, begin the usual way by considering a one-dimensional concentration profile and by finding the rate of change of number of Bicoid molecules in the box at position x by considering the flux into $(J_m(x - \Delta x/2))$ and out of $((J_m(x + \Delta x/2))$ the box using arguments like those made in class. However, you need to generalize that treatment by accounting for the fact that a Bicoid molecule has the probability $r\Delta t$ of degrading in time interval Δt, where $r \approx 1/\tau$, where τ is the degradation time.

(C) Now solve this equation in steady-state by finding the general solution subject to the boundary condition that $J(0,t) = j_0$ and $J(L,t) = 0$. Make sure you explain what these boundary conditions mean relative to the biology of the problem. Suggest approximations that can be made to simplify the result, specifically, can you exploit the fact that the embryo is much larger than the decay length to simplify the solution?

(D) Describe the observed concentration profile of Bicoid along the anterior-posterior axis of the fly mathematically. What is the functional form? Experimentally, Thomas Gregor has found that the Bcd profile is an exponential of the form $Bcd(x) = Bcd_0 e^{-x/\lambda}$, does that jibe with your solution?

(E) The paper by Drocco *et al.* uses a photoactivatable fluorescent protein to measure the lifetime of the Bicoid protein. Read the paper (available on the course website) and explain the technique in one paragraph. You might find it useful to draw a schematic plot such as shown in Figure 1f of the paper.

(F) What is the value of the decay constant λ for the gradient shown in Figure 10? To estimate this magnitude, you can just fit "by eye" by plotting your solution for different values of Bcd_0 and λ. Now, compare the measured λ value with that you can predict by plugging in realistic values of D, τ into your solution. To make this possible, read the papers by Abu-Arish *et al.* and Drocco *et al.*, provided on the course website.

(G) One of the most important and interesting ideas to come out of the idea of positional information contained in morphogen gradients was the so-called

French flag model which we will explore here. This model posits that the Bicoid concentration dictates the position of the cephalic furrow. As seen in Figure 12, the idea of the model is that boundaries in the embryo are determined by threshold values of the morphogen. The idea of the model is that if the gene dosage gets changed, as seen in the mutant profile, the boundary will still occur at the same value of the morphogen. That hypothesis is enough to determine the shift in boundary position with gene dosage.

To test this model, we will analyze several experiments (Nusslein-Vohlhard and Driever and Liu *et al.*) where they measured cephalic furrow position as a function of different dosages of the *bicoid* gene in embryos. An exponential gradient of Bicoid is described by

$$Bcd(x, \lambda, \alpha, Bcd_0) = Bcd_0\, \alpha\, e^{-x/\lambda}, \tag{6}$$

where x is the position along the embryo, Bcd_0 is the Bicoid concentration at $x = 0$, λ is the decay constant of the gradient and α is the Bicoid dosage, with $\alpha = 1$ corresponding to the wild-type. Work out a model for the position of the cephalic furrow x_{new} as a function of the gene dosage α, the morphogen gradient decay length λ and the position of the wild-type cephalic furrow, x_{CF}.

(H) Note that, given a measured $x_{CF} \approx 32\%$ of the embryo length, your model has no free parameters. Compare the prediction from your model with the data for x_{new} vs. α obtained by Nusslein-Vohlhard, and by Driever and Liu *et al.*. Comment on how well your prediction matches the data that is provided with the homework. What could be going on?

5. The Protein-mRNA Ratio.

In this problem we go beyond the calculation on mRNA production we did in class, and think about how transcription and translation shape the protein-to-mRNA ratio inside cells.

(A) In class, we described the temporal evolution of the number of mRNA molecules using the equation

$$m(t + \Delta t) = m(t) + r_m \Delta t - \gamma_m m(t) \Delta t. \tag{7}$$

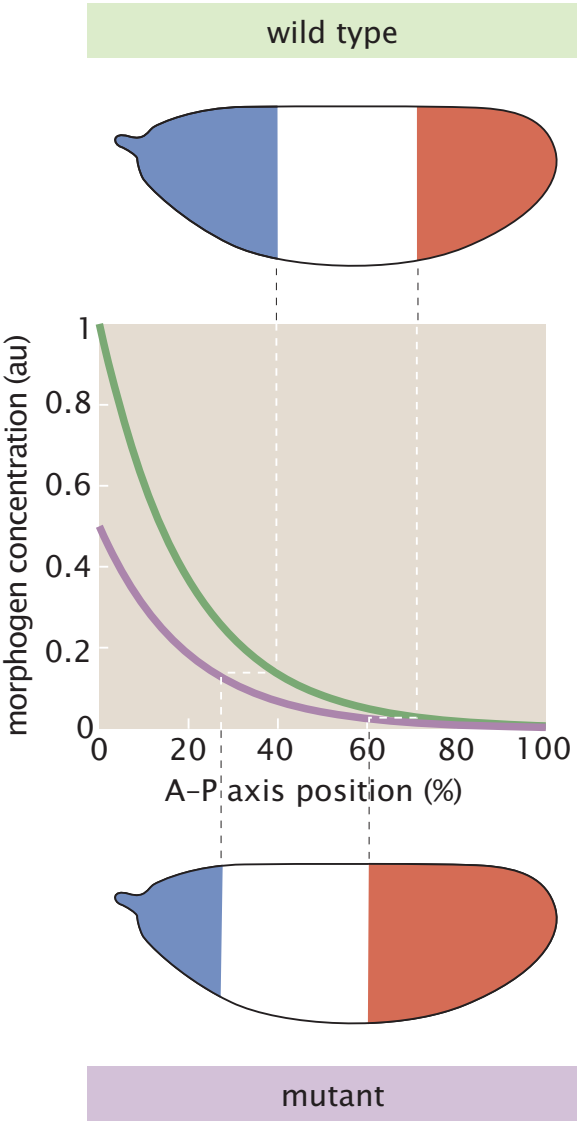


Figure 12: Concept of the French flag model.

Here, $m(t)$ is the number of mRNA at time t, r_m is the rate of mRNA production, and γ_m is the mRNA decay rate. Write the corresponding equation for the number of protein molecules given a rate of protein production *per mRNA* of r_p and a protein decay rate γ_p. Make sure to incorporate the fact that the number of mRNA molecules present will determine how many proteins are produced in a time interval Δt.

(B) Calculate the ratio of protein to mRNA in steady state, p_{SS}/m_{SS} and show that it is given by r_p/γ_p. Find typical values for the various model parameters in *E. coli* and estimate the ratio of proteins to mRNA molecules. How do your numbers compare to those measured in Figure 3C of Taniguchi *et al.*, which is provided on the course website?

(C) Using flies with different dosages of Bicoid-GFP, Petkova *et al.* measured the relation between the number of *bicoid* mRNA molecules deposited by the mother, and the resulting number of Bicoid proteins. Read their paper (available on the course website) and make sure you understand how their Figure 3 is generated. Assuming that Bicoid-GFP is in steady state, use Figure 3 from Petkova *et al.* to estimate the ratio r_p/γ_p. Use the value for the degradation rate obtained by Drocco *et al.* discussed in Problem 3 in order to calculate r_p.

N. Your Turn.

In this final problem, I want you to construct a thoughtful syllabus for how you would teach a course on Physical Biology. You have ten weeks, two classes of 90 minutes each per week. Make sure to give a sense of whether your homeworks will involve computation, whether you will give an exam, etc. But more importantly, what is the content? What do you want students to leave the course with? What are the top five skills you want them to leave with? What are the top five insights you want them to leave with? You have 20 lectures, so I want to hear what each and every lecture will be about. How much powerpoint? How many calculations on the blackboard. For this problem, send a pdf (nothing but pdf accepted and zero credit for stuff like powerpoint or word files) to Vahe and Rob.