BIOLOGICAL PHYSICS - Part III MAJOR OPTION - QUESTION SHEET 2019 - will be covered by 2 supervisions - written answers will be circulated after material is covered

MODULE A

Lecture 1: Intro and concepts

A1 **Intro**

(a) Roughly 2-3 kg of bacteria are in your large intestine. Estimate the number of bacteria that this corresponds to. Estimate the number of human cells in your body, and compare the two figures.
(b) Look up estimates of protein density in a cell. Estimate cell volumes for bacteria, yeast and a typical mammalian cell. Work out rough total number of proteins in these cells.
(c) Look up rough fractions of protein, RNA, lipids and other classes of molecules in the dry mass of cells. What is a typical water content in a cell?
Explore, and keep in mind, the online resource:
http://bionumbers.hms.harvard.edu/

- The number of bugs is about 10^{14}, human cells about $4\,10^{13}$, so there are roughly 3 bacteria cells per human cell (other estimates give this ration as up to 10). About 3% of the body mass is bacteria. There are about 1000 species in our gut.
- bug volume: $2\,\mu m^3$. Yeast volume: $50\,\mu m^3$. Mammalian cell volume: $2000\,\mu m^3$.
  Proteins have density of between 1.3 and 1.4. Water weight fractions have been measured from 60% in yeast, up to 77% in some mammalian cells. So dry mass per volume is between 300 and 500 mg/ml, and average density of a cell is about 1.1.
  A typical protein has 350 amino acids, and mass roughly 35 kDa, which is $6\,10^{-26}$ kg.
  So if we take 30% as dry mass (and of this about 55% proteins) in cells, then number of proteins is:
  bugs: $30/100 * 55/100 * 2\,10^{-18} * 1.1\,10^3 / (6\,10^{-23}) = 5 * 10^6$.
  yeast: 10^8, mammalian: $5\,10^9$.
- needed for above.

A2 **Growth laws - a numerical exercise on growth dynamics** *(adapted from BBSRC Sysmic course.)*
A modified growth model, with similar behaviour to logistic growth, can be developed from the differential equation:

$$dN/dt = r(N)N,$$

where

$$r(N) = r_0\left(1 - \left(\frac{N}{K}\right)^2\right).$$

For the following use parameters values $r_0 = 0.0347$ and $K = 1000$.
i) Use Matlab or your favorite package to plot dN/dt.
ii) Use this plot to infer how the population will change at different values of N.
iii) Describe how this plot differs from the equivalent plot for the logistic growth model.
iv) Using Matlab (or otherwise!) find the value of N at which the growth rate is a maximum.
(Hint: there are two solutions but we require $N > 0$.)
v) Find the solution of this differential equation, for example using Matlab's dsolve command.
vi) Create two figures overlaying plots showing population growth, for the model introduced here and for Logistic growth. Do this for both $N_0 = 100$ and $N_0 = 5000$.
vii) Describe the differences between the graphs.
make a .m file with the script below, and call it from the Matlab prompts window. (this is uploaded as `bug_colony_2.m`)

```
close all\\
dL = @(L,r0,K) r0*L*(1-(L/K)^2  );\\
fun = @(L) dL(L,0.0347,1000);\\
limits = [0 1200];\\
fplot(fun,limits)\\
line( [0,1200], [0,0],'color','k')\\
ylabel('dL/dt')\\
xlabel('L')\\
syms L r0 K\\
f = r0*L*(1-(L/K)^2);\\
df = diff(f,L);\\
ddf=diff(df);\\
mdf = matlabFunction(df,'vars',[L,r0,K])\\
mddf = matlabFunction(ddf,'vars',[L,r0,K])\\
solve(ddf==0, 'L')\\
```

```
fun = @(L) mdf(L,0.0347,1000);\\
limits = [0 1200];\\
fplot(fun,limits)\\
line( [0,1200], [0,0],'Color','k')\\
ylabel('d/dL  (dL/dt)')\\
xlabel('L')\\
syms L r0 K t L0
L=dsolve('DL=r0*L*(1-(L/K)^2)', 'L(0)=L0', 't')
L=simplify(L)
mL = matlabFunction(L,'vars',[L0,K,r0,t])

funL100 = @(t) mL(100,1000,0.0347,t);
funL5000 = @(t) mL(5000,1000,0.0347,t);

figure(2)
limits = [0 200];
fplot(funL100,limits,'Color','r')
hold on
fplot(funL5000,limits,'Color','r')
%line( [0,1200], [0,0],'Color','k')
ylabel('L (red) and N (blue)')
xlabel('t')


syms N r0 K t N0
N=dsolve('DN=r0*N*(1-(N/K))', 'N(0)=N0', 't')
N=simplify(N)
mN = matlabFunction(N,'vars',[N0,K,r0,t])

funN100 = @(t) mN(100,1000,0.0347,t);
funN5000 = @(t) mN(5000,1000,0.0347,t);
fplot(funN100,limits,'Color','b')
hold on
fplot(funN5000,limits,'Color','b')
%line( [0,1200], [0,0],'Color','k')
```

Running this script makes a figure in which L is the modified logistic, and N is the logistic model. The square term makes the approach to steady state much faster.
run it as:

```
EDU>> bug_colony_2
```

giving

```
mdf =

    @(L,r0,K)-r0.*(1.0./K.^2.*L.^2-1.0)-1.0./K.^2.*L.^2.*r0.*2.0


mddf =

    @(L,r0,K)1.0./K.^2.*L.^2.*-3.0+1.0


ans =

  (3^(1/2)*K)/3
 -(3^(1/2)*K)/3
```

These are the two values of L that give maximum growth; we look at the positive solution. K=1000, so this evaluates to 577.3503. Matches with the plot.

```
L =
(-K^2/(exp(2*K^2*(log(-(K^2 - L0^2)/L0^2)/(2*K^2) - (r0*t)/K^2)) - 1))^(1/2)

L =
```

```
((K^2*L0^2*exp(2*r0*t))/(K^2 - L0^2 + L0^2*exp(2*r0*t)))^(1/2)

mL =
    @(L0,K,r0,t)sqrt((K.^2.*L0.^2.*exp(r0.*t.*2.0))./(K.^2-L0.^2+L0.^2.*exp(r0.*t.*2.0)))

N =
K/(exp(K*(log((K - N0)/N0)/K - (r0*t)/K)) + 1)

N =
(K*N0*exp(r0*t))/(K - N0 + N0*exp(r0*t))

mN =
    @(N0,K,r0,t)(K.*N0.*exp(r0.*t))./(K-N0+N0.*exp(r0.*t))
```

A3 Cell size control

Consider the growth control models discussed in the first lecture as examples of "physics" based models (in this case models based not on stat mech as many others in the course, but trying to tease out consistent mechanisms from data). The "added size" to a cell is given by a $s_b + \Delta$, where s_b is the size at birth, $-1 < a < 1$ is a parameter that distinguishes 'timer' control ($a = 1$), from 'adder' ($a = 0$) and 'sizer' ($a = -1$). $\Delta > 0$ is a fixed added size.
Show that in an organism performing symmetric divisions, starting from a cell of arbitrary size s_b, the size of its progeny will eventually converge. Find an expression for how this steady state size depends on Δ and a.
Advanced: Make some considerations on the speed of convergence (in generations) for varying a.
For thought: What experiments are done to study these control mechanisms? Size control is necessary for all populations of dividing cells; do you think the control mechanism will be universal? What might you imagine as being plausible 'timer' and 'sizer' mechanisms at a molecular level?

Answer: write down what happens from one generation to next, and (lucky!) we'll spot that a well defined series shows up.

$$s_b \to s_b + a s_b + \Delta$$

$$\frac{(a+1)s_b + \Delta}{2} \to (a+1)\left[\frac{(a+1)s_b + \Delta}{2}\right] + \Delta = \frac{a+1}{2}((a+1)s_b + \Delta) + \Delta$$

$$(a+1)\left[\frac{(a+1)\left[\frac{(a+1)s_b+\Delta}{2}\right] + \Delta}{2}\right] + \Delta = \frac{a+1}{2}\left[\frac{a+1}{2}[(a+1)s_b + \Delta]\right] + \Delta$$

...

$$\Delta + \frac{a+1}{2}\Delta + \left(\frac{a+1}{2}\right)^2 \Delta + \left(\frac{a+1}{2}\right)^3 \Delta + \ldots + \left(\frac{a+1}{2}\right)^n \Delta + \left(\frac{a+1}{2}\right)^n a s_b$$

$$= \sum_{i=0}^{n} \frac{a+1}{2}^i \Delta + \left(\frac{a+1}{2}\right)^n a s_b$$

$$= \Delta \frac{1 - \left(\frac{a+1}{2}\right)^{n+1}}{1 - \frac{a+1}{2}} + \left(\frac{a+1}{2}\right)^n a s_b = 2 s_b$$

drop all power of n terms because small, gives:

$$\left(1 - \frac{a+1}{2}\right) s_b = \Delta/2$$

so

$$s_b = \frac{\Delta}{1-a}.$$

For convergence: my first approach was to add a small ϵ to the growth size, and see how quickly it got diluted. I took the ratio of the sizes at successive generations and it seems the ϵ term cancels out. The ratio is $1/(a+1)$ which means that with $a = 1$ (timer) we have the slowest convergence, and with $a \to -1$ the fastest.

Lecture 2: Numbers in the cell, and Central Dogma

A4 Small numbers - (2.7 from [Phillips et al., 2013])

In considering the cell to cell variability, one aspect is the random partitioning of molecules to daughter cells. Derive the result that if there are N molecules, and this partitioning follows purely a random process, then the variance in the number n_1 of the molecules that go to cell 1 is:
$\langle n_1^2\rangle - \langle n_1\rangle^2 = N\,p\,q$, where p is the probability of the molecule going into daughter cell 1, and $q = 1 - p$ is the probability of that molecule going to cell 2.
If the molecule of interest are fluorescent, and the mother cell has total fluorescence I_{tot}, show that you can expect the variability in the fluorescence of daughter cells (I_1 and I_2) to follow: $\langle (I_1 - I_2)^2\rangle = \alpha I_{tot}$, where α is a coefficient relating the number of molecules to the fluorescence as $I = \alpha N$.
For thought: what would you think if there was found to be deviations from this last results in experiments? Can you think of plausible complications from purely the experimental side? Can you think or do a quick literature search for possible biological effects?
Numerical investigation
Can you modify the 'countingbydilution.m' script to plot the distributions of $\langle (I_1 - I_2)^2\rangle$ for some values of I_{tot} ? Do these resemble more closely Gaussian distributions (normals) or log-normal?

The number of ways we can divide N indistinguishable molecules into two cells is given by:

$$W = \frac{N!}{n_1!(N-n_1)!},$$

where N is a total number of molecules and n_1 is the number of molecules that went to daughter cell 1. The probability of a particular sequence of N molecules going to daughter cell 1 and daughter cell 2, where the probability of the molecule going to daughter cell 1 is p and to daughter cell 2 is q, is

$$p^{n_1} q^{N-n_1},$$

since if n_1 molecules went to daughter cell 1, $N - n_1$ molecules had to go to cell 2. Thus, the probability distribution is given by

$$P(n_1, N) = \frac{N!}{n_1!(N-n_1)!} p^{n_1} q^{N-n_1}.$$

To get the average number of molecules going to daughter cell 1, we have to sum over all the configurations. In particular, we have

$$\langle n_1\rangle = p\frac{\partial}{\partial p}\sum_{n_1=0}^{N} \frac{N!}{n_1!(N-n_1)!} p^{n_1} q^{N-n_1}.$$

We can calculate the second moment of the distribution similarly as

$$\langle n_1^2\rangle = p\frac{\partial}{\partial p} p\frac{\partial}{\partial p}\sum_{n_1=0}^{N} \frac{N!}{n_1!(N-n_1)!} p^{n_1} q^{N-n_1}$$

Now, we can calculate the difference between the second and the square of the first moment of the distribution as:

$$\langle n_1^2\rangle - \langle n_1\rangle^2 = \left(p\frac{\partial}{\partial p}\left(p\frac{\partial}{\partial p}\sum_{n_1=0}^{N} P\right)\right) - \left(p\frac{\partial}{\partial p}\sum_{n_1=0}^{N} P\right)^2.$$

To rewrite this in a more useful form, we recall the binomial theorem, namely

$$(p+q)^N = \sum_{n_1=0}^{N} \frac{N!}{n_1!(N-n_1)!} p^{n_1} q^{N-n_1}$$

and we get

$$\begin{aligned}
\langle n_1^2\rangle - \langle n_1\rangle^2 &= p\frac{\partial}{\partial p}\left(p\frac{\partial}{\partial p}(p+q)^N\right) - \left(p\frac{\partial}{\partial p}(p+q)^N\right)^2 \\
&= p\frac{\partial}{\partial p}\left(pN(p+q)^{N-1}\right) - (pN(p+q)^{N-1})^2 \\
&= pN(p+q)^{N-1} + p^2N(N-1)(p+q)^{N-2} - p^2N^2(p+q)^{2N-2}.
\end{aligned}$$

Now we use the fact that $p + q = 1$ resulting in

$$\begin{aligned}
\langle n_1^2\rangle - \langle n_1\rangle^2 &= pN + p^2N(N-1) - p^2N^2 \\
&= pN + p^2N^2 - p^2N - p^2N^2 \\
&= pN - p^2N = pN(1-p) = Npq
\end{aligned}$$

This can now be used to examine the relation between the intensities of the two daughter cells and their mother. First, we observe that

$$\langle (I_1 - I_2)^2 \rangle = \langle (2I_1 - I_{tot})^2 \rangle,$$

where we have used the fact that $I_1 + I_2 = I_{tot}$. If we expand the binomial and use the relation $I = \alpha N$, this implies

$$\langle (I_1 - I_2)^2 \rangle = 4\alpha^2 \langle N_1^2 \rangle - 4I_{tot}\alpha \langle N_1 \rangle + I_{tot}^2.$$

From the previous part of the problem, for the case in which $p = q = 1/2$, we have

$$\langle N_1 \rangle = N/2$$

and

$$\langle N_1^2 \rangle = \frac{N}{4} + \frac{N^2}{4}.$$

As a result, we have

$$\langle (I_1 - I_2)^2 \rangle = \alpha^2 N = \alpha I_{tot},$$

as we set out to prove. This observation is the basis of a very clever scheme for counting proteins.

A5 On timing of DNA replication, speed of forks - (3.3 from [Phillips et al., 2013]).

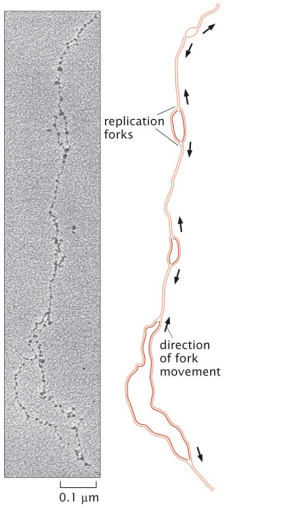


Figure 3.35 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

The electron microscopy image, and the helpful schematic cartoon (both from [Alberts et al., 1994]) show a snapshot in time of replication forks duplicating the genome of a drosophila fly.
(a) Estimate the fraction of the total fly genome captured in the micrograph. (The total fly genome is about 1.8×10^8 nucleotide pairs in size.)
(b) Extrapolating from this snapshot, estimate the number of DNA polymerase molecules in a eukaryotic cell like this one from drosophila.
(c) There are 8 forks in this micrograph. Estimate the lengths of the strands between the forks 4 and 5 counting from the bottom. If a fork moves at a replication speed of about 40 bp/s how long will it take forks 4 and 5 to collide?
(d) Given the mean spacing between the forks seen here, estimate how long it will take to replicate the entire genome.

(a) By using the scale bar and measuring the length of the picture, we conclude that this segment of the genome is roughly $1.4\,\mu$m in length. This corresponds to

$$\#\text{base pairs} = \frac{1400\,nm}{0.34nm/\text{bp}} \simeq 4000\text{base pairs}.$$

As a result this corresponds to a fraction of the genome of roughly

$$f = \frac{4000\,\text{bp}}{2 \times 10^8 \text{bp}} \simeq 2 \times 10^{-5}.$$

(b) The crudest estimate is to imagine that all of the DNA polymerase molecules are bound to the DNA. In this figure, there are 8 such polymerases, one for each replication fork. Hence, the estimated total number of polymerases is

$$\#polymerases \simeq 8 \times \frac{2 \times 10^8}{4000} \simeq 4 \times 10^5.$$

(c) The distance between replication forks 4 and 5 is roughly $0.2\,\mu$m or 200 nm. This corresponds to roughly 600 bp. If each polymerase is moving at a rate of roughly 100 bp/s this implies that it will take only 3 seconds for the two forks to collide.
(d) Note that this figure is from the replication in an early embryo in the fly. At this stage in its development, the cell cycle time is of order ten minutes so the cells are replicating and dividing very quickly. With the assumptions we have made about the constant density of replication forks (probably not true over the entire cell cycle as evidenced by the different lengths of the bubbles of replicated DNA) and the 100 nt/s replication rate, our estimates lead us to the conclusion that the replication time has a time scale of roughly five seconds since the amount of DNA already copied is roughly half of the total. Note that this estimate is troubling in that it makes replication appear implausibly fast. A search of the literature leads to the conclusion that the polymerase may not be moving as fast as 100 nt/s. In addition, the density of replication forks may not be as uniformly high as shown in the figure.

(TURN OVER

A6 **Ribosomes - (adapted from 3.4d from [Phillips et al., 2013])**

E. coli is growing in a condition where cells divide every 3000 s. If each cell contains typically 20000 ribosomes, and assuming no ribosome degradation, how many RNA polymerase molecules must be synthesising rRNA (RNA that codes for ribosomes) at any instant? What percentage of RNA polymerase molecules in E. coli are involved in transcribing rRNA genes?

During the course of a cell cycle, the cell must double the numbers of all its parts. Therefore if there are 20,000 ribosomes and the division time is 3000 seconds, the cell must produce 20,000/3000 = 6.7 ribosomes per second, at least. This operon codes for the 16S and 23S ribosomal subunits, which are made directly from the RNA transcript, and each ribosome contains only one of each of these. This means that the rate of ribosome production is at most equal to the rate of transcription of this operon times the number of copies of this operon.

The speed of transcription by RNAP is between 0.04 and 0.05 kb/sec, so roughly 50 nucleotides/s.

Human rRNA is 7216 nt, E.Coli can't find number... But this means 144 s to make one transcript for a ribosome (made of various subunits). (Phillips uses 120 s for a rRNA transcript in E.Coli.) So it would require 120 transcribing RNAP to produce roughly one transcript per second. Therefore to make 6.7 transcripts per second requires $6.7 \times 120 \simeq 800$ RNAP to be always transcribing the operon.

This (literature search) is roughly half of all the RNAP in the cell.

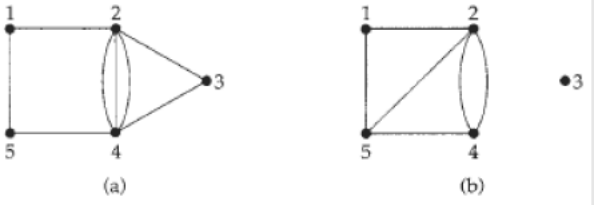
Further, these numbers clearly point to the existence of more than one copy of the rRNA operon. Indeed, as the reader can see by consulting the rrnDB (the Ribosomal RNA Operon Copy Number Database), the average number of copies of this operon in E. coli is roughly seven.

Lecture 3: Networks

A7 **General definitions**

Let G be a graph with n vertices labelled $1, 2, 3, ..., n$. The adjacency matrix $A(G)$ of G is the $n \times n$ matrix in which the entry in row i and column j is the number of edges joining the vertices i and j.

(a) *Write down the adjacency matrix of each of the following graphs:*



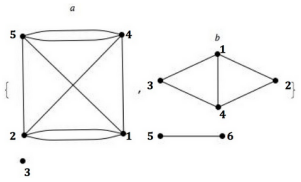
$$(a)\begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 3 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 3 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$(b)\begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

(b) *Draw the graph represented by each of the following adjacency matrices:*

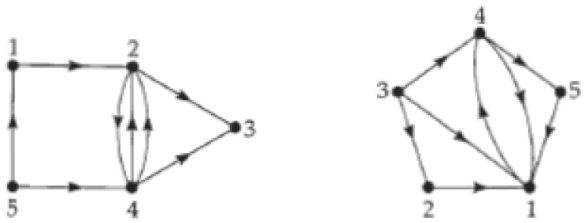
$$\begin{pmatrix} 0 & 2 & 0 & 1 & 1 \\ 2 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 2 \\ 1 & 1 & 0 & 2 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$



Let D be a digraph (directed graph) with n vertices labelled $1, 2, 3, ..., n$. *The adjacency matrix* $A(D)$ *of D is the* $n \times n$ *matrix in which the entry in row i and column j is the number of arcs (directed "arrows") from vertex i to vertex j.*

(c) *Write down the adjacency matrix of each of the following digraphs:*



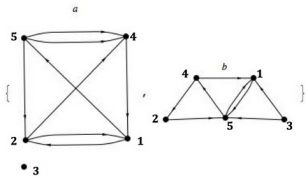
$$(a)\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$(b)\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

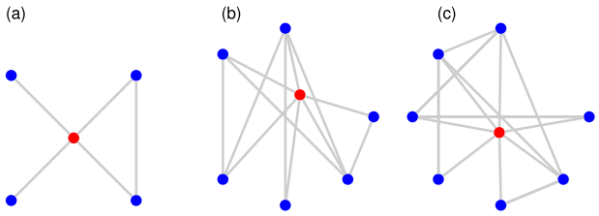
(d) *Draw the digraph represented by each of the following adjacency matrices:*

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$



(e) *Calculate the clustering coefficients for the red nodes in the graphs below:*



Convince yourself that the maximum number of edges between k vertices in an undirected graph with no loops is $k(k-1)/2$.

Think: How is this related to a special property of the adjacency matrix?

Definition: The clustering coefficient of a vertex u, denoted by C_u, is given by the ratio of the number of edges between its nearest neighbours, e, and the maximum possible, given by $k(k-1)/2$ where k is the degree of u. Thus

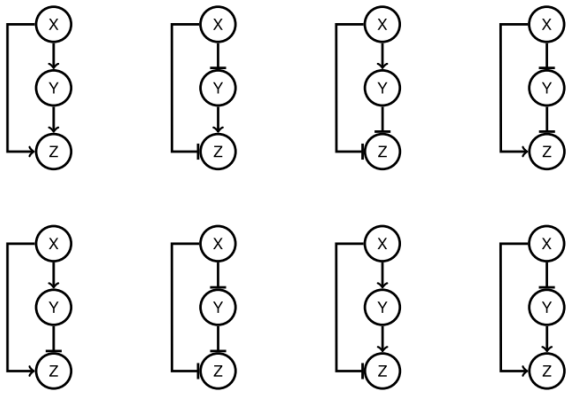
$$C_u = \frac{2e}{k(k-1)}.$$

The red node in graph (a) below has a degree of 4 (and therefore 4 nearest neighbours) so $k = 4$. There is only one edge between these 4 nearest neighbours which means $e = 1$. Therefore the clustering coefficient for this node is 1/6 or 0.1667. The red node in graph (b) has a degree of 6 with 6 edges between its neighbours and so the clustering coefficient is 0.4. In graph (c) the red node has a clustering coefficient of 0.33.

The adjacency matrix for a graph with k vertices has a maximum of k^2 elements. For a graph with no loops, the elements on the main diagonal should be zero, thus the maximum number of elements would be $k \times k - k = k(k-1)$. Considering that the adjacency matrix of an undirected graph is symmetrical about the main diagonal, it can be concluded that each edge has been represented twice in the matrix. Thus, the maximum number of edges would be $k(k-1)/2$.

An interesting quantity is the average clustering coefficient over all the nodes in the network: It tells us how likely the neighbours of a graph are themselves connected. We can also define the clustering distribution in a very similar way to the degree distribution.

(f) *Classify each of the feedforward motifs below as coherent or incoherent (see glossary in notes handout):*



Top are coherent, bottom are incoherent.

A8 **Numerical network question. (problems 4 and 5, p.219 of [Sneppen and Zocchi, 2005].)**

Simulate a network that grows with preferential attachment: say that at each time step either with probability $R_{new} = 0.1$ *one node with one link is added, or else a link between two existing nodes is added with a preferential attachment to each node.*

Let the network reach a steady state, by removing nodes plus all their links with small probability (try $\epsilon \simeq 10^{-3}$ *or* $\simeq 10^{-4}$*. Quantify the degree distribution. (Optional: Quantify the degree of correlation of the number of links per node, between nearest neighbour nodes. Is this same or different to that of a random graph?)*

Then, let $n(k,t)$ *be the number of nodes with connectivity k at time t. Can you find the analytical expression for the steady-state distribution of* $n(k)$ *for different values of* R_{new}*?*

The master equation that describes the change in number of nodes with degree k over time is

$$\frac{dn(k,t)}{dt} = R_{new}\frac{(k-1)n(k-1,t) - kn(k,t)}{\sum_k kn(k,t)} + (1-R_{new})\frac{2(k-1)n(k-1,t) - 2kn(k,t)}{\sum_k kn(k,t)},$$

where the first term represents the probability that one node is added to the network and connected to one of the existing nodes, and the second term corresponds to the probability that two existing nodes are connected by an extra link.
Since at every time step I add a node, $\sum kn(k) = 2E = 2t$, and also the number of nodes are being added proportionally to time. Then,

$$\frac{dn(k,t)}{dt} = \frac{n(k,t)}{t} = (R_{new}-2)\frac{kn(k,t) - (k-1)n(k-1,t)}{2t} = \frac{R_{new}-2}{2t}\frac{d(k\cdot n)}{dk} = \frac{R_{new}-2}{2t}(n + \frac{dn}{dk}k)$$

By separating variables and integrating, one then obtains:

$$n(k) \propto k^{\frac{4-R_{new}}{R_{new}-2}}$$

A9 Thinking about size of regulatory networks, from [Sneppen and Zocchi, 2005]

Regulatory genetic networks are essential for epigenetics and thus for multicellular life, but are not essential for life. In fact, there exist prokaryotes with nearly no genetic regulation. The course handout shows a figure for the fraction of regulators as a function of genome size, for a number of prokaryotic organisms. Understand the following:

- *If life was just a bunch of independent switches, i.e. if living cells could be understood as composed of a number of modules (genes regulated together) each, for example, associated with a response to a corresponding external situation, then the fraction of regulators would be independent of the number of genes N.*
- *If life was simply hierarchical, with each gene controlling a certain number of downstream genes, then again the number of regulators would grow linearly with N.*
- *If life would have been controlled using the maximum capacity for combinatoric control, even fewer transcription factors would be necessary. Show that if the regulation of each gene could include all transcription regulators available, then in principle the total state of all genes could be specified with only* $\log_2(N)$ *regulators.*

Last point: To see this, imagine that one had 10 regulators that can be on or off. Then one can, in principle, specify 2^{10} different states, and thus specify all possible states for up to $N = 2^{10}$ genes.

MODULE B

Lectures 4-6: Evolution: general principles, neutral models, selection, and the role of space

B1 Accumulation of mutations in our guts [Phillips et al., 2013]

(a) Suppose that a very particular combination of k point mutations is required for a pathogenic E. coli strain to arise in the human gut and that all these mutations must arise in one cell division (as could be the case if each subset of these mutations are deleterious). If the point mutation rate per base pair per cell division is μ, what is the probability m_k *that this occurs in a single cell division? (Assume that the probabilities of different mutations are independent)*
(b) In a human large intestine, the density of bacteria is estimated to be about $10^{11.5}$ *per milliliter, of which a fraction of about* 10^{-4} *are E. coli. Estimate how many E. coli per person this implies (mean large intestine volume is approximately 13 mL). In a population of N humans, with n E. coli in each of their guts, in T generations of the E. coli, estimate the total probability* p_k *that the particular combinations of k mutations occurs at least once.*
(c) With the population of Cambridge over one year, what are the chances this occurs for k = 2? And k = 3? Keep in mind that μ is between 10^{-10} *and* 10^{-9} *mutations per base pair per generation in E. coli. Standard lab strains of E. coli divide every 20 minutes. A low-end estimate of the division rate of E. coli in human guts is about once every few days. Why is this more realistic? Given these and other uncertainties, how big are the uncertainties in your estimate.*

- The probabilities π_i that mutation i will occur are independent and we do not care whether other mutations are also there. Then $m_k = \Pi_{i=1}^{k}\pi_i = \mu^k$
- $p_k = m_k nNT$

B2 Protein mutation rates [Phillips et al., 2013]

Random mutations lead to aminoacid substitutions in proteins that are described by the Poisson probability distribution $p_s(t)$*. Namely, the probability that s substitutions at a given amino-acid position in a protein occur over an evolutionary time t is*

$$p_s(t) = \frac{e^{-\lambda t}(\lambda t)^s}{s!}$$

where λ is the rate of aminoacid substitutions per site per unit of time. For example, some proteins, like fibrinopeptides evolve rapidly (λ = 9 substitutions per site per 10^9 *years). Histones, on the other hand, evolve slowly with λ = 0.01 substitutions per* 10^9 *years.*

(a) What is the probability that a fibrinopeptide has no mutations at a given site in 1 billion years? What is this probability for a histone?
(b) We want to compute the average number of mutations $\langle s\rangle$ over time t,

$$\langle s\rangle = \sum_{s=0}^{\infty} s p_s(t).$$

First, using the fact that probabilities must sum to 1, compute the sum $\sigma = \sum_{s=0}^{\infty}(\lambda t)^s/s!$. Then, write an expression for $\langle s\rangle$, making use of the identity

$$\sum_{s=0}^{\infty} s\frac{(\lambda t)^s}{s!} = (\lambda t)\sum_{s=1}^{\infty}\frac{(\lambda t)^{s-1}}{(s-1)!} = \lambda t\sigma$$

(c) Using your answer in (b), determine the ratio of the expected number of mutations in a fibrinopeptide to that of a histone, $\langle s\rangle_F/\langle s\rangle_H$.

- $p_0 = \exp(-\lambda t)$. That is $\exp(-9)$ for fibrinopeptides and $\exp(-0.01)$ for histones.
- $1 = \sum_s \frac{e^{-\lambda t}(\lambda t)^s}{s!} = e^{-\lambda t}\sum_s \frac{(\lambda t)^s}{s!}$, therefore, $\sigma = e^{\lambda t}$. $\langle s\rangle = \sum_s s\frac{e^{-\lambda t}(\lambda t)^s}{s!} = e^{-\lambda t}\lambda t\sigma = \lambda t$

B3 Genetic drift and heterozigosity

(a) The heterozigosity H of a population is the probability that two random individuals will carry different alleles. Write an expression for the heterozigosity for a biallelic population of size N, where one allele is at frequency X.
(b) Using your favorite scripting language, track the heterozigosity over time of three biallelic populations of constant size $N = 10$,50 and 100, where allele a is carried by 10% of the initial population and allele A by the rest 90%. Use the average over 100 simulations. Implement the simulations using the following different ways:
(i) Use a Wright-Fisher model by sampling each generation using a binomial distribution.
(ii) Use a Moran model, where at each step a random individual is picked to replicate and one to die (the two individuals might be the same).
(iii) A random-walk model, where at each step the frequency of allele a increases of decreases by a step $1/N$ with 50% probability.
What is the decay of H as a function of population size N? How does it differ between the different ways of simulating the process?

- $H = 2X(1-X)$ from binomial sampling.
- In the WF model, $H \propto e^{-t/N}$, while in the Moran model $H \propto e^{-2t/N^2}$. The diffusion model gives a linear decay in time with slope $2/N^2$.

B4 Mutation-selection balance

Suppose a haploid population is characterized by two alleles A (wild-type) and a (mutants). Mutations continuously transform allele A to a at a rate μ. However, selection continuously purges mutants since they grow at a rate $1-s$ compared to the wild-type which grows at rate of 1.
(a) Write the system of differential equations that describe the dynamics of the two alleles over time. What is the equilibrium frequency of allele a? (assume most individuals carry allele A and that reverse mutations are negligible).
(b) How does the equilibrium frequency change if we now have a diploid population where only the homozygous aa are less fit growing at a rate $1-s$?

The dynamic is defined by mutation and selection. Remember that in a biallelic population with alleles p and q, the allele dynamic follows $p'(t) = \frac{w_1-w_2}{\overline{w}}pq$ where w_i are the relative fitness coefficients. Then, in this case

$$\begin{cases} A' = sAa - \mu A \\ a' = -saA + \mu A \end{cases}$$

where we approximate $\overline{w} = A + (1-s)a = 1$ if $a \ll A$. Then, at equilibrium $a_{eq} = \mu/s$.
In the diploid case, individuals can be of genotype AA, aa and Aa. Given the frequencies A and a for the two alleles, the three genotypes will have frequencies $p_{AA} = A^2$, $p_{aA} = 2aA$ and $p_{aa} = a^2$. The mean fitness of the population will be $\overline{w} = A^2 + 2aA + (1-s)a^2 = 1 - sa^2$. The change in frequency for A due to selection and random mating after one generation will be

$$A(t+1) = \frac{A^2(t) + \frac{1}{2}2a(t)A(t)}{1-sa^2(t)} = \frac{A(t)[A(t)+a(t)]}{1-sa^2(t)} = \frac{A(t)}{1-sa^2(t)}$$

Then, we can write its dynamics by sending the time interval to zero

$$A' = A(t+1) - A(t) = \frac{sA(t)a^2(t)}{1-sa^2(t)}$$

and therefore the dynamics of $a = 1 - A$ is $a' = -\frac{sAa^2}{1-sa^2}$ due to selection and random mating. By adding mutations, we find

$$a' = -\frac{sAa^2}{1-sa^2} + \mu A$$

Then, at equilibrium, assuming $a \ll A$, we find $a_{eq} = \sqrt{\mu/s}$.

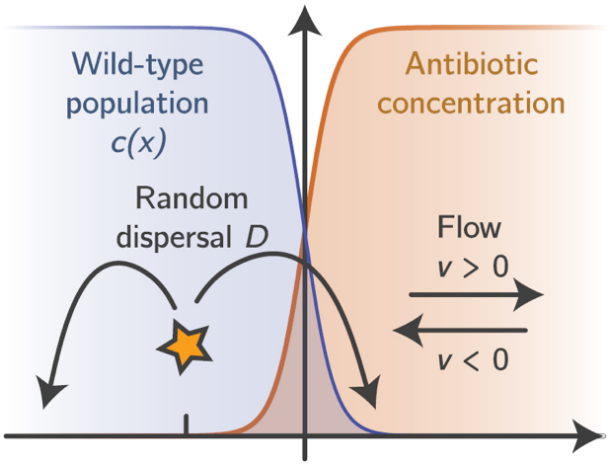
B5 **Bak-Sneppen model [Sneppen, 2014]** *Simulate a Bak-Sneppen model for 100 species placed along a line in a variant of the model where only one of the neighbors is updated at each step. Plot the selected $B_{\min}$ as a function of time, as well as the maximum of all previous $B_{\min}$s. How do the minima of B change as time progresses toward a steady-state (look at the envelope defined by the maximum over all $B_{\min}$ at earlier times)?*

B6 **Emergence of resistance in antibiotic gradient** *A bacterial population resides at the edge of an antibiotic gradient. The wild-type population $c(x,t)$ replicates at rate $a(x) = 1$ and dies at a rate $b(x) = 2\Theta(x)$ where $\Theta(x)$ represents the Heavy-side function, so that its dynamics can be written as*

$$\partial_t c = \partial_x^2 c + c(1-c) - b(x)c.$$

Random mutations can occur in the wild-type population and confer resistance to the antibiotic so that the mutant m will grow at a rate $a(x) = 1$ and die at a rate $b(x) = 0$ and its dynamics is described by

$$\partial_t m = \partial_x^2 m + m(1-c).$$



(a) If $u_x(t)$ represents the probability distribution that a mutation born at position x will survive until time t, use a birth-death process to derive the following equation describing the dynamics of u

$$\partial_t u = \partial_x^2 u + [1 - c(x)]u - u^2.$$

(b) Assuming that $c(x)$ at steady-state can be approximated by a step function that mirros the antibiotic gradient $c(x) = \Theta(-x)$, derive the probability distribution $u(x)$ by solving the equation above for steady-state using the mechanical analogy of a particle in a potential (note that $x < 0$ and $x > 0$ have to be solved separately).

If $u_x(t)$ represents the probability that a mutation that arose at position x has survived until time t, the probability that it will survive an extra time ϵ is given by the branching process that sums the probabilities of survival of at least two lineages if the lineage replicates or that nothing happens plus diffusion:

$$u_x(t+\epsilon) = \epsilon a(x)\left\{1 - [1-u_x(t)]^2\right\} + \{1 - \epsilon\,[a(x)+b(x)]\}\,u_x(t) + \epsilon D\,\{u_{x+\delta x}(t) + u_{x-\delta x}(t) - 2u_x(t)\}\,.$$

By sending ϵ and δ_x to 0, we get the general differential equation

$$\partial_t u = \partial_x^2 u + [a(x) - b(x)]u - a(x)u^2$$

that we have seen in class. Now, $a(x) = 1$ and $b(x) = c(x)$ due to the logistic growth of the mutants (no additional explicit death rate).

If assume that $c(x) = \Theta(-x)$, then the steady-state equation to solve is

$$\begin{cases} \partial_x^2 u - u^2 = 0 \text{ for } x < 0 \\ \partial_x^2 u + u - u^2 = 0 \text{ for } x > 0 \end{cases}$$

For $x < 0$, the corresponding potential is $U(u) = -u^3/3$ and the total energy $E = 0$. Therefore, $u(x)$ is given by

$$\frac{du}{dx} = \sqrt{2u^3/3}.$$

By separating variables, one obtains

$$u(x) = \left(\frac{1}{\sqrt{u_0}} - \frac{x}{\sqrt{6}} \right)^{-2}$$

where $u_0 = u(0)$, which we will determine by continuity.
For $x > 0$, $U(u) = u^2/2 - u^3/3$ and $E = 1/6$. The expression of $u(x)$ is given by

$$\frac{du}{dx} = \sqrt{1/3 - u^2 + 2u^3/3}.$$

You can use Mathematica to perform the integral and obtain:

$$u(x) = \frac{3}{2} \tanh\left[\frac{x + \hat{x}}{2} \right]^2 - 1/2.$$

By imposing that $u(x)$ is continuous and differentiable at $x = 0$, one finds that $u_0 = 1/\sqrt{3}$ and $\hat{x} = 2\text{arctanh}(\frac{1}{3}\sqrt{3 + 2\sqrt{6}})$

MODULE C

Lectures 8 and 9: Bioelectricity: sensing and neural biophysics

C1 **Ions and channels in an algae cell** The table below gives some data for a marine alga. The extracellular fluid is seawater; the 'plasmalemma' is the outer cell membrane separating the outside of the cell from the cytoplasm; the 'tonoplast membrane' separates the cytoplasm from an interior organelle, the vacuole.
(a) The table gives some of the Nernst potentials across the two membranes. Fill in the missing ones.
(b) The table does not list the charge density of impermeant macroions in the cytoplasm. What is their charge density, in mM?
(c) The actual measured membrane potential across the tonoplast membrane is +76 mV. Which ion(s) must be actively pumped across the tonoplast membrane and in which direction(s) for this to be the case?
(d) Suppose we selectively shut down the ion pumps in the tonoplast membrane, but the cell metabolism continues to maintain the listed concentrations in the cytoplasm. The system then relaxes to an equilibrium across the tonoplast membrane. What will be the approximate ion concentrations inside the vacuole, and what will be the final equilibrium potential?

Ion	Vacuole (mM)	Cytoplasm (mM)	Extracellular (mM)	v^{nernst} (plasmalemma) (mV)	v^{nernst} (tonoplast) (mV)
K^+	530	425	10	?	-5.5
Na^+	56	50	490	+57	?
Cl^-	620	30	573	-74	+76

C2 **Instability of membrane potential** Consider a membrane with two types of channel only. One is a 'leak' channel with conductance g_{leak} independent of voltage and with Nernst potential equal to zero, and the other opening in response to a increasing voltage, with a Nernst potential V_r. Then a dynamical model very similar to the one in the lecture approximates the dynamics of the system as:

$$\begin{aligned} C\frac{dV}{dt} &= -G_{leak}V - gNp_{open}(V - V_r) \\ \frac{dp_{open}}{dt} &= -\frac{1}{\tau(V)}\left[p_{open} - p_{open}^{eq}\right]. \end{aligned} \tag{1}$$

Recall the form of p_{open} from the lectures.
Consider the steady states conditions of the system above, which give two functions $f = f(V)$.
Plot these on the same axes for the numerical values: $V^* = 70\,\text{mV}$, $\beta q = 0.1\,(\text{mV})^{-1}$, G_{leak}/gN=0.1. Observe for how many values of voltage these two functions cross.
Crossings are special conditions where both dynamical equations are at steady state. Can you determine which of these points are stable or unstable?
Can you interpret this result as a memory? as an amplifier of an input signal?

C3 **Axon signal in Matlab** Code the time dependent Cable equation and solve it using an ODE solver in Matlab. Investigate from the numerical solution the speed of propagation.
Extension: code up the dynamics of channel open/closed/inactive states and reproduce spike shapes.

C4 **Leaky Integrate and Fire Model** In its simplest form, a neuron can be modeled as a "leaky integrator" of its input current $I(t)$:

$$\tau_m \frac{dv}{dt} = -V(t) + RI(t),$$

where $V(t)$ represents the membrane potential at time t, τ_m is the membrane time constant and R is the membrane resistance. This equation describes a simple resistor-capacitor (RC) circuit where the leakage term is due to the resistor and the integration of $I(t)$ is due to the capacitor that is in parallel to the resistor. The spiking events are not explicitly modeled in the LIF model. Instead, when the membrane potential $V(t)$ reaches a certain threshold V^{th} (spiking threshold), it is instantaneously reset to a lower value V_r (reset potential) and the leaky integration process described by the above starts anew, with the initial value V_r.
(a) Show that in the case of constant input and $V_r = 0$ the solution to this equation is

$$V(t) \;=\; RI(1 - \exp(-t/\tau_m.$$

(b) Assuming the systems starts with $V(0) = 0$, find the time of the first spike t_1 and the time between successive spikes. How does the mean firing rate depend on input current?

MODULE D

Lecture 10-12: Molecular motors

D1 **Velocity of thermal ratchet** Suppose we have M identical S-ratchets with bolts length L apart, and we want to find the probability distribution $P(x, t)$ given the potential energy used in class (each bolt stores potential energy ϵ and there is a load f on each ratchet). It helps to imagine the ratchet as a circle, where bolts are always reset. In this case, the solution is time-independent and periodic.
(a) Use the same arguments we used in class (consider small discrete intervals δx) to derive an equation for the net number of ratchets crossing a given point in space per unit time (j^{1D}) and then the Smoluchowski equation. Show that the time-independent solution in equilibrium is the Boltzmann distribution.
(b) To go beyond equilibrium, show that the function

$$P(x) = C(b\exp[-(x - L)f/k_BT] - 1)$$

solves the Smoluchowski equation with potential energy $U_{tot} = fx$. Verify that j^{1D} is constant.
(c) Now, by considering the continuity of the solution across piece-wise continuous potentials, fix b and note that C cancels out using the equations you have found.
(d) By considering a small time interval Δt find the velocity of the thermal ratchet. Show that, in the perfect ratchet limit ($\epsilon \gg k_BT$)

$$v = \left(\frac{fl}{k_BT}\right)^2 \frac{D}{L}(exp[fL/k_BT] - 1 - fL/k_BT)^{-1}.$$

Plot this and comment what this means in the context of biochemical reactions such as protein translocation across membranes.

D2 **Driven diffusion in pore translocation** Consider a newly-synthesized RNA molecule of length L passing through a small pore in the nuclear membrane from the nucleus to the cytosol. Suppose that as soon as the leading end of the molecule passes through the pore, a protein binds irreversibly to it, thereby preventing the molecule from sliding back through the pore to the nucleus. Consider a constant friction ζ on the RNA molecule, principally caused by interactions of the RNA with the tight pore.
(a) Estimate how long it takes the RNA to diffuse into the cytosol in the absence of a membrane potential.
(b) If there is an electrical potential difference V across the membrane, and the RNA has charge per unit length σ, show that the mean translocation is

$$t = \frac{\zeta}{\sigma V}\left(L - \frac{k_BT}{\sigma V}\left[1 - \exp\left(-\frac{\sigma VL}{k_BT}\right)\right]\right).$$

(c) Now suppose tht the membrane potential is zero, but further proteins in the cytosol bind irreversibly to the RNA with constant rate per unit length α. Show that the translocation time varies as

$$t \approx \left(\frac{\zeta}{k_BT}\right)^{2/3} \alpha^{-1/3}L.$$

D3 **Equilibrium polymerization** A single-stranded protein filament is made up of monomers, which bind and detach from the filament ends with equilibrium constant K. If the equilibrium concentration of filaments of lenth n is c_n, show that

$$a_n = a_1^n = \exp(-n/n_0),$$

where $a_n = c_n/K$ and $n_0 = -1/\ln a_1$. Hence show that the average length of the polymer filaments (excluding monomers) is

$$\langle n\rangle = 1 + \frac{1}{1 - a_1}.$$

In the case where the total concentration of protein subunits c_t is large, such that $a_t \gg 1$, show that the concentration of unpolymerized monomers is given by

$$a_1 \approx 1 - \frac{1}{\sqrt{a_t}}$$

and the mean filament length is

$$\langle n \rangle = 1 + \sqrt{a_t}.$$

D4 **Dynamic instability of microtubules** Microtubules alternate between phases of growth at velocity v_+, and shrinkage at velocity v_-. Switches from growth to shrinkage are caused by catastrophes, which occur randomly at rate f_{+-}. Rescue from the shrinking to the growing phase occurs randomly at rate f_{-+}. Show that, in the steady-state, the fraction of microtubules with length n is

$$p(n) = n_0^{-1} \exp(-n/n_0),$$

where

$$n_0 = \frac{v_- v_+}{v_- f_{+-} - v_+ f_{-+}}$$

provided that $v_- f_{+-} > v_+ f_{-+}$.
How is the distribution of microtubule lengths regulated in the cell? What useful role does the dynamic instability of microtubes play in the cell?

D5 **Bacterial flagellar motor** In an experiment, the bacterial flagellar motor is found to rotate through 10 revolutions in a mean time $\langle t_{10} \rangle = 11.9$s with standard deviation $\sigma(t_{10}) = 0.19$s. According to one model, the rotor turns by making a large number of small steps, with each step occurring at a constant rate and moving the rotor through a fixed angular displacement. If one assumes that the stochastic nature of the stepping accounts for all of the variance in the experimental data, how many steps per revolution does the data imply? How does this correlate with the observation that approximately 800 protons traverse the membrane for each revolution of the motor? (The membrane potential in *E. coli* is 150 mV). Comment on the observed value of torque required to stop the motor from turning, $G_{stall} = 3 \cdot 10^{-18}$ Nm.

MODULE E

Lecture 13 and 14: Biopatterning

E1 **One component Turing pattern**

(a) Write the reaction-diffusion equation given in your lecture notes for the case of a single morphogen whose concentration within the cells is Y_r. Then consider a small perturbation of the uniform steady-state $Y_r = Y^*$ of the form $Y_r = Y^* + y(t)e^{i(2\pi r)/\lambda}$ and derive the dynamical equations for the amplitude $y(t)$.
(b) Assuming that there are N cells in the system and that they are arranged in a ring so that $r = 1$ cells has the $r = 2$ and $r = N$ cells as its nearest neighbors, what are allowed values of the wavelength λ for the periodic perturbation?
(c) Derive the conditions under which the uniform steady-state is unstable to a small periodic perturbation. Does this one-component Turing system lead to a spatially periodic pattern of morphogen concentration?

E2 **Wavelength of Turing Patterns** Consider the following continuum reaction-diffusion equations

$$\begin{aligned} \partial_t X &= D_x \partial_x^2 X + f(X, Y) \\ \partial_t Y &= D_y \partial_x^2 Y + g(X, Y) \end{aligned}$$

for morphogen concentrations X and Y, which are both functions of position. Introducting the following notation:

$$a_{11} = \frac{\partial f}{\partial X}|_{(h,k)}; a_{12} = \frac{\partial f}{\partial Y}|_{(h,k)}; a_{21} = \frac{\partial g}{\partial X}|_{(h,k)}; a_{22} = \frac{\partial g}{\partial Y}|_{(h,k)}$$

where h and k are steady-state values of X and Y. We are interested in the stability of the homogeneous steady-state $(X, Y) = (h, k)$ with respect to a perturbation that is periodic in space, since such instability suggests the spontaneous formation of a pattern of morphogens.
(a) Linearize the reaction-diffusion equations in the vicinity of the fixed point $(X, Y) = (h, k)$. Consider the time evolution of a periodic perturbation of the form $(\delta X, \delta Y) = (\eta(t)e^{iqx}, \epsilon(t)e^{iqx})$ in the vicinity of the fixed point (h, k), following the same strategy we used in class. In particular, derive expressions for the eigenvalues of the rate matrix $\boldsymbol{A}$ (analogous to that defined in class) in terms of its determinant and trace.
(b) Show that a necessary condition for linear stability of a fixed point in the absence of diffusion is that at least one of a_{11} and a_{22} is negative.
(c) Show that for a fixed point that is linearly stable in the absence of diffusion, a necessary condition for stability in the presence of diffusion is that exactly one of a_{11} and a_{22} is negative.
(d) Assuming that $a_{22} < 0$, show that at a Turing bifurcation, the characteristic length of the emerging pattern from the homogeneous steady-state scales like $\lambda^{-2} \sim \lambda_x^{-2} - \lambda_y^{-2}$, where $\lambda_x = \sqrt{D_x/a_{11}}$ and $\lambda_y = \sqrt{-D_y/a_{22}}$.
(e) Comment on the physical implications of these results. What is the wavelength of an oscillatory instability?

E3 **Notch-Delta Model** The Notch-Delta concept was discussed in class. This question looks at the limit when the decay rate of Notch is much greater than that of Delta.

(a) Reason that in the limit of $\nu = \gamma_D/\gamma_N \ll 1$, the Notch activity of the two cells quickly settles into a steady-state. What are the resulting dynamical equations for the evolution of Delta that follow from the equations given in class in this limit?
(b) Using the functional forms for the Notch and Delta activation rates, $F(D)$ and $G(N)$, described in class, obtain a phase portrait for the dynamics of Delta in the two cells. Based on your phase portrait, show that in the long-term limit, the system will settle into a steady state in which one cell assumes the primary fate while the other assumes the secondary fate.

MODULE F

Lecture 15: Protein production

F1 **On protein production ODE**

Show by integration of

$$\frac{dc_p(t)}{dt} = k - \gamma c_p(t)$$

that the time-dependent protein concentration $c_p(t)$ is given by:

$$c_p(t) = \frac{k}{\gamma}(1 - \mathrm{e}^{-\gamma t}) + c_p(0)\mathrm{e}^{-\gamma t}.$$

Make sure you understand what all the parameters mean in this expression.
Consider now a protein which activates its own production, i.e. the protein is only produced when a protein molecule is bound to the promoter region.
(a) write down a differential equation for the dynamics of protein production in this case.
(b) Solve for the steady state protein concentration $c_P^{(ss)}$. What interesting feature do you notice about your solution?

Lecture 16: Biochemical Noise

F2 **Master equation**

The notes give expressions for chemical reactions for a model of protein production in which mRNA is produced from a source, protein is produced from mRNA, and both protein and mRNA are degraded to a sink.
(a) Write down the chemical master equation for the probability $p(N_P, N_M, t)$ of having N_P protein molecules and N_M mRNA molecules at time t for this model, supposing that the rate of mRNA production is k, the rate of mRNA degradation is γ, the protein production per mRNA molecule is s, and the rate of protein degradation is μ.
(b) Explain why the steady-state solution of this master equation must obey the relation

$$\sum_{N_p} p^{(ss)}(N_P, N_M) = \frac{1}{N_M!}\left(\frac{k}{\gamma}\right)^{N_M} \mathrm{e}^{-\frac{k}{\gamma}}.$$

Lectures 17 and 18: Regulation of Gene Expression

F3 **Polymerase binding** (based on Q6.3 p277 [Phillips et al., 2013])

The probability of promoter occupancy can be computed both by statistical mechanics or by using equilibrium constants.
(a) Write an expression for the probability of finding RNA polymerase bound to the promoter as a function of the equilibrium constants for specific and nonspecific binding.
(b) *In vitro*, the dissociation constant of RNA polymerase binding to nonspecific DNA is approximately 10 μM, and the dissociation constants of RNA polymerase to the *lac P1* and T7A1 promoters are 550 nM and 3 nM respectively. Use these constants and the result from (a) to estimate *in vivo* binding energies of RNA polymerase to *lac P1* and T7A1 promoters.

F4 **On activators/repressors** (based on Q19.1 p889 [Phillips et al., 2013])

Repression is a quantitative measure of the reduction in the level of gene expression due to the action of a repressor molecule. For the simplest model of repression, make a plot comparing repression in the two limits of a weak and a strong promoter. Show that in the case of a strong promoter, (not so for weak promoter), the repression depends on the number of polymerase molecules in the cell.

F5 **The lac Operon** (based on Q19.2 p889 [Phillips et al., 2013])

Using the data in figure 19.22 of [Phillips et al., 2013] (these three relevant figures are on slide 21 of this module, or thereabout.), determine the in vivo binding energy of Lac repressor to each one of its operators, and reproduce Figure 19.23.
Use the result just obtained, and the data on repression in cells with two operators present (which leads to DNA looping) to determine the looping energy, and reproduce the curves on Figure 19.27.

F6 **On regulation** (based on Q19.3 p889 [Phillips et al., 2013]).

The 'sensitivity' in gene regulation is defined as the steepness of the transition (e.g, in activation from the 'off' to the 'on' state) in response to change in number of transcription factors. It can be measured as the slope on a log-log plot in the power law region near the transition.
Using the stat-mech models discussed in lectures, determine how the sensitivity depends on the relevant parameters for the following regulatory motifs in the case of a weak promoter:
(a) simple activation
(b) simple repression

(c) two binding sites where the same species of repressor can bind. They can recruit each other and repress NA polymerase independently. What happens if the interaction is turned off ? (assume that the sites have same binding energy).

F7 **Optional extended question** - Build a partition function for a new case: work through Q19.4 p889 [Phillips et al., 2013].

MODULE G

Lecture 20: Stochastic reactions

G1 **Question on Gillespie algorithm**

(a) Download and run GillespieAlgorithm.m . Make sure you understand the logic of the algorithm, the implementation, and the figures that are being output.

(b) Plot the bias of production/decay event as a function of time. Explain intuitively what is happening.

(c) Plot the timestep as a function of time.

(d) Can you edit the GillespieAlgorithm.m file to study the (simpler) system $A + B \rightarrow C$ for varying initial numbers of $N_A = N_B$? (Take care in how the production rate should be modified compared to the GillespieAlgorithm.m example.)

Lecture 21: General elements of dynamical systems

G2 **General concepts and types of stabilities**

What is a non-linear system ?
What types of fixed points may systems with one or two variables have, and how may one determine which fixed points are stable?

G3 **Phase portraits**

Plot the phase space for the logistic growth model.

G4 **Numerical investigation of Lotka-Volterra.**

Work out nullclines, phase portrait and time dependence plots.

Lecture 22: Genetic switches and oscillators

G5 **On switches and oscillators** (based on Q19.13 p891 [Phillips et al., 2013])

Consider a very simple genetic switch, obtained when an activator activates its own production.
(a) Write a rate equation for the time evolution of the activator. Include a term for a basal rate of production even in the absence of activator. (note: you may want as a first step to make a table of 'states', 'weights' and 'rates', analogous to Figure 19.45 of [Phillips et al., 2013], for this case.)

G6 **Numerical investigation of a toggle system** A system with switch behaviour can be made with two repressor proteins, u and v:

$$\begin{aligned} \frac{du}{dt} &= \frac{\alpha_1}{1+v^\beta} - u \\ \frac{dv}{dt} &= \frac{\alpha_2}{1+u^\gamma} - v \end{aligned} \tag{2}$$

Let's work on it with Matlab. We have defined a function we call toggle, in which some numerical values are set. Download toggle.m from the course website.
Use the numerical integration ode45 as follows:
[t,y] = ode45(@toggle,[0 10],[2 3]);
to integrate in time between 0 and 10, with initial conditions $u = 2$, $v = 3$. The command plot(t,y) plots the concentrations of both u and v over the specified time range.
Adapt the code above to calculate and plot over the time interval $0 - 15$, for the two initial conditions $0.1, 1$ and $5, 4$. Think: what is common between these two solutions? What has changed? Think about what is happening, both mathematically, and "biologically".

G7 **Numerical investigation of repressilator** The repressilator network of three mRNA species m_1, m_2, m_3 and the three protein species p_1, p_2, p_3 can be modeled by the system of equations:

$$\begin{aligned} \frac{dm_1}{dt} &= -m_1 + \frac{\alpha}{1+p_3^n} + \alpha_0 \qquad \frac{dp_1}{dt} = -\beta(p_1 - m_1) \\ \frac{dm_2}{dt} &= -m_2 + \frac{\alpha}{1+p_1^n} + \alpha_0 \qquad \frac{dp_2}{dt} = -\beta(p_2 - m_2) \\ \frac{dm_3}{dt} &= -m_3 + \frac{\alpha}{1+p_2^n} + \alpha_0 \qquad \frac{dp_3}{dt} = -\beta(p_3 - m_3), \end{aligned}$$

which can be coded as in the file repressilator.m (download it, and understand the short code).

We can get time series and a phase space plot for the time interval $t = [0, 15]$, and initial conditions
$m_1, p_1, m_2, p_2, m_3, p_3 = 0, 1, 0, 1, 0, 1$ with the commands in the script rep_plots.m (download and run):
Note that the trajectory is not straight and curves, or spirals around the steady state. Now try initial conditions $[0, 1, 0, 2, 0, 5]$.
Something quite exciting should happen now...

References

[Alberts et al., 1994] Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. (1994). *Molecular Biology of the Cell*. Garland Publishing, New York.

[Phillips et al., 2013] Phillips, R., Kondev, J., Theriot, J., Garcia, H., and Orme, N. (2013). *Physical Biology of the Cell, 2nd Ed.* Garland Science, London.

[Sneppen, 2014] Sneppen, K. (2014). *Models of Life*. Cambridge University Press, Cambridge.

[Sneppen and Zocchi, 2005] Sneppen, K. and Zocchi, G. (2005). *Physics in Molecular Biology*. Cambridge University Press, Cambridge.