PH419: Physics of Biological Systems
Assignment 2

1. **Segmentation patterns in Drosophila embryos (Concept: Numbers and scales)**

   The embryonic development of the Drosophila embryo serves as a testbed for developmental biology. In the initial stages of development, the embryo undergoes synctial division , in which the nuclei divide without forming associated cellular membranes. This process continues until the thirteenth nuclear cycle, and at this stage, all nuclei share a common cytoplasm, and material can diffuse throughout the embryo.

   The drosophila embryo can be modeled as a spherocylinder, with a length of $500\mu m$ for the length of the cylindrical region, and a radius of $100\mu m$. During the first to the ninth cycles, the nuclei are in the bulk of the embryo. After the end of the ninth nuclear cycle, most of the nuclei migrate to the surface from the bulk. Given that there are approximately 6000 nuclei on the surface of the embryo after the 13th cycle, what is the fraction of nuclei that migrated to the surface at the ninth cycle? Also calculate the areal density of nuclei at the surface at the end of the 13th cycle, i.e. the number of nuclei per unit area.

   Shown below is the pattern of gene expression at the end of the 13th cycle for three genes, Bicoid, Even-skipped, and Caudal. As can be seen from the figure, the genes formorganize in a segmented pattern, which is the precursor to heterogenous anatomical features of the adult fly.

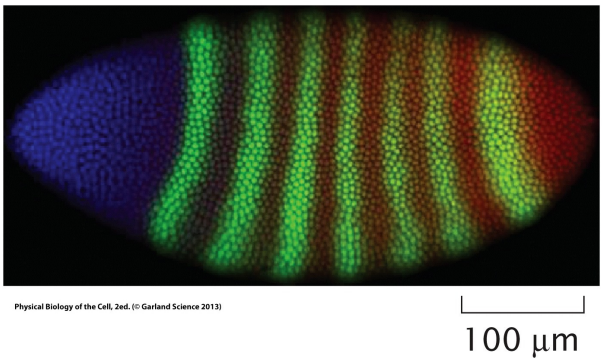

Figure 1: Pattern of gene expression in Drosophila embryo after 13 nuclear cycles. Bicoid is labeled in blue, Even-skipped in green, and Caudal in red.

   From the figure, estimate the width of the green segmented regions, and thence calculate the number of nuclei, $n_{segment}$ in each of these regions. This provides an estimate of the number of nuclei that show distinct patterns of gene expression.

2. **Partitioning of carboxysomes (Concept: Active vs Random partitioning)**

We discussed in class that partitioning of organelles and proteins from a mother cell to the two daughter cells can often be understood in terms of random partitioning, resulting in a binomial distribution.

We now consider an experiment that determines the partioning of carboxysomes in cyanobacteria. Carboxysomes are capsidlike protein microcompartments that are responsible for carbon fixation (conversion of carbon dioxide to organic compounds). The experiment quantifies the partitioning of carboxysomes from a mother cell to the daughter cells. (Ref.: Spatially Ordered Dynamics of the Bacterial Carbon Fixation Machinery, Savage *et. al.*, Science, 327, 1258, 2010)

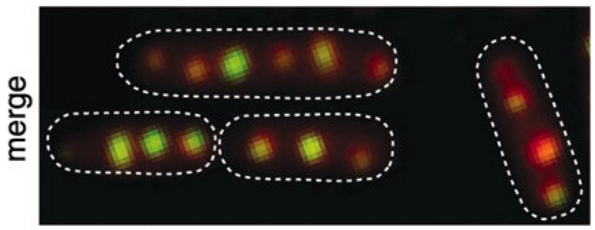

Figure 2: Fluoroscently labeled carboxysomes in mother and daughter cells

Cells with exactly six carboxysomes were examined during cell division, and the number of carboxysomes delivered to one daughter was recorded. The experimental data is summarized in the table below,

Number of carbxysomes in daughter cells	Observed frequency
0	0
1	0
2	0.147
3	0.706
4	0.147
5	0
6	0

Compare the experimental distribution with the theoretical binomial distribution, using a histogram plot, and verify that random partioning is not an effective model in this case.

Experimentally it is observed that carboxysomes are arranged in an ordered fashion, with roughly equal spacing along the cell length. The measured diffusion constant

of carboxysomes is abnormally low, with $D = 4.58 \times 10^{-5} \mu m^2/s$. This implies that carboxysomes in the left half of the mother cell are more likely to go to daughter cell 1, while those in the right half are more likely to go to daughter cell 2.

We can now attempt to construct an active partitioning model. For a cell with six carboxysomes, let us subdivide the carboxysomes into two classes, L and R. The three in the left half of the mother cell are of type L, while the three in the right half are of type R. The L type carboxysomes go to daughter cell 1 with probability p, and to daughter cell 2 with probability $q = 1 - p$, with $p \gg 0.5$. The R type carboxysomes go to daughter cell 2 with probability p, and to daughter cell 1 with probability q. Now estimate the probability of having daughter cells with $2, 3$ and 4 carboxysomes. Is there a value of p which reproduces the experimental data?

If not, can you construct a better model of active partitioning the fits the observed data?

3. **Open Reading Frames (Concept:Probability)**

Assume that the nucleotides A, G, T, C occur with equal probability (and independently) along a segment of DNA.

(a) In the genetic code, a stop codon (or termination codon) is a nucleotide triplet within messenger RNA that signals a termination of translation into proteins. The stop codon can be encoded by either of the following three triplets, (i) TAG, (ii) TAA, and (iii) TGA. Calculate the probability p_s that a randomly chosen triplet of bases corresponds to a stop signal.

(b) An open reading frame (ORF) is the part of a reading frame that has the potential to be translated. Thus an ORF of length N is defined as a sequence of N non-stop *triplets* followed by a stop codon. What is the probability for an ORF of length N?

(c) The genome of *E. coli* has roughly 5×10^6 bases per strand, and is in the form of a closed loop. There are six possible reading frames, i.e. six possible ways to read a linear sequence of bases appearing in *E. coli*. Justify why one has six possible reading frames.

(d) If the bases in the *E. coli* genome were arranged completely randomly, how many ORFs of length 600 (a typical protein size) would be expected on the basis of chance?

To compute the actual distribution of ORFs in E. coli you will need to download the complete sequence of its genome. The sequence file is provided with this assignment (U00096.fna). Download this sequence file. Further information about the sequence can be found at https://www.ncbi.nlm.nih.gov/genome/167

(e) Write a program that goes through all consecutive (non-overlapping) triplets looking for stop codons. Record the distance L between consecutive stop codons. Repeat this computation for the 3 different reading frames (0, +1, +2) in this direction.

(f) Plot the distribution for the ORF lengths L calculated above, and compare it to that for random sequences.

(g) Estimate a cut-off value L_{cut} , above which the ORFs are statistically significant, i.e. the number of observed ORFs with $L > L_{cut}$ is much greater than expected by chance.

4. **One-dimensional FRAP (Concept: Diffusion)**

We discussed in class how FRAP experiments can be used to determine the diffusion constant of various proteins.

Shown below are the FRAP recovery curves from two such experiments in spinach leaves, *Spinacia olearacea*, which studies the diffusion of a chlorophyll-protein complex on a patch of an isolated spinach grana membrane. Fig. 3a corresponds to experiments on a native grana membrane patch, while Fig. 3b corresponds to grana membrane-liposome fusion product. The figures are taken from the following reference, Protein Diffusion and Macromolecular Crowding in Thylakoid Membranes, Kirchoff *et.al.*, Plant Physiol. 146, 1571, 2008

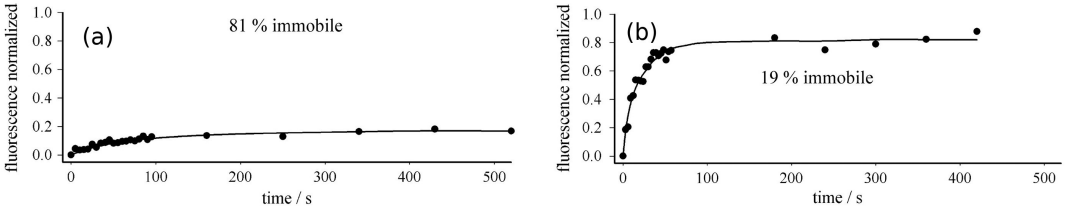

Figure 3: FRAP recovery curves for (a) native grana membranes; and (b) grana membrane-liposome fusion product

The data files for the two figures are provided with this assignment (frapdata1.dat and frapdata2.dat). Assume that the data can be analysed using the one-dimensional FRAP framework discussed in class. Assuming a linear length of $5\mu m$ and a photobleached region in the center of length $1\mu m$ (from experimental data), calculate the diffusion constant D of the chlorophyll-protein complex in the two cases. Note that in the experiments, a fraction of the protein is immobile and does not diffuse. In which case does the protein have a larger diffusivity?.

From the reference given above, note down the experimental values of the diffusion constant in the two cases. How do they compare with the values you obtain?

5. **Two-dimensional FRAP (Concept: Diffusion)**

   The goal of this problem is to generalize the one-dimensional treatment of FRAP done in class. Consider a cell as a planar circle of radius R uniformly covered with freely diffusing fluorescent proteins. Imagine that the laser photobleaches a hole of radius a in the middle of the cell. Note that we ignore the presence of the nucleus. Work out the concentration of fluorescent proteins in the cell as a function of position and time in analogy with the one-dimensional treatment of the problem done in the chapter. Compute the number of molecules in the hole after photobleaching as a function of time.