MCB137L/237L: Physical Biology of the Cell
Spring 2022
Homework 9:
(Due 4/7/22 at 11:00am)

Hernan G. Garcia

"Biology is catching up" - PAM Dirac

NOTE FOR MCB237L STUDENTS: You don't need to do problem (1). Please do problem (3) and then choose to do one of the following problems: (2), (4) or (5). You can do more problems for extra credit (15% of total possible score per exxtra problem).

1. A Feeling for the Numbers in Biology: Round two

Now that you have written your first estimate vignette in the style of *Cell Biology by the Numbers*, it's time to get ready for the second estimate! You can choose to write a vignette or to present a poster the last day of classes. Your estimate can be on a completely new topic with respect to your first estimate, or it can build on your previous estimate.

Write a short paragraph describing the estimate you're interested in writing about. Note that the objective at this point is not for you to have a finished estimate, but to have an outline of the calculation you plan to do so that we can give you feedback. Send this paragraph as an email to Hernan, Yasemin and Yovan by **4/7 (when Homework 9 is due)**.

2. Ligand-receptor problem from the perspective of statistical mechanics

In Homework 6, we calculated ligand-receptor binding from the perspective of rate equations and dissociation constants. In this problem we explore a third route to compute the probability of a receptor being bound by a ligand based on statistical mechanics. Note that you can complement this problems by reading the paper "A First Exposure to Statistical Mechanics for Life Scientists: Applications to Binding" (Garcia2007b) on the course website.

(a) Imitate the statistical mechanics protocol given in class by showing the states, energies, multiplicities and weights for a lattice consisting of Ω lattice sites and L ligands. Find an expression for p_{bound} in terms of the difference in energy of a ligand when in solution, ε_{sol} and the energy when the ligand is bound to the receptor, ε_b.

(b) Consider that the lattice sites in our lattice model (Figure 1 of Homework 6) have size v and hence that the concentration is $[L] = L/\Omega v$ and use that insight to arrive at an expression for the dissociation constant in terms of the microscopic parameters ε_{sol} and ε_b. Do this by comparing the results of this part of the problem with your result from Homework 6.

3. Mutation per generation in humans

Comparing genetic sequences has served as a useful tool for determining how various organisms are related to each other. With the advent of the "genomic era," we no longer have to infer how living organisms are related to each other based on morphological traits alone. In this problem, we will begin to get a sense of the time scales over which mutations accumulate in genetic sequences and how we can use these mutations as a molecular clocks for determining the relationships between various organisms.

In this problem, we are ultimately interested in estimating the total number of mutations that are passed on in each human generation. As a first step, we must estimate the number of mutations that accumulate in a single cell division.

(a) Given that the human genome is 3 billion basepairs long and is replicated with an incredible fidelity of only one error in every 10^{10} basepairs per replication, how many mutations do you expect to see after one genome duplication?

With this number of mutations per genome duplication in hand, we can next tackle how many mutations are passed on by a mother and a father. Recall that while many mutations may occur in a given human, only those that accumulate in the gametes (egg and sperm) will actually be passed on. To determine the number of mutations that we expect to be passed on, we will need to consider the formation of the egg and the sperm separately as males and females have different developmental pathways regarding gametogenesis (see Figure 2).

As a primer for thinking about gametogenesis, let's briefly review the difference between mitosis and meiosis. Mitosis is the process by which a somatic cell duplicates its genome and then divides into two cells. Thus in a human, mitosis yields two cells with 46 chromosomes each. Meiosis, however, is the process by which a cell duplicates its genome and then proceeds to undergo two cell divisions, ultimately resulting in four cells with 23 chromosomes. This means that each round of mitosis requires one genome duplication and each round of meiosis requires one genome duplication (despite having two cell divisions).

In humans, females are born with all of their eggs nearly fully developed and they produce no new egg cells throughout the rest of their life. As illustrated in the top half of Figure 2, every developed egg is the result of 22 rounds of mitosis and 1 round of meiosis, yielding a total of 23 genome replications. This means that every egg a woman produces has undergone 23 genome replications regardless of a woman's age.

(b) Given the 23 genome duplications that occur in the process of forming an egg, how many mutations do you expect a woman to pass on to her children?

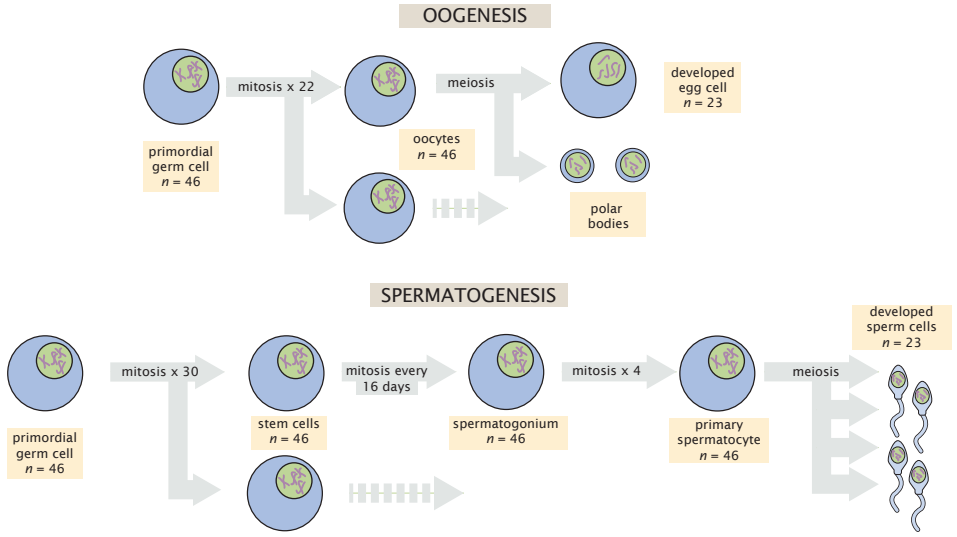

Figure 1: Schematic of oogenesis and spermatogenesis in humans. n refers to the number of chromosomes, where somatic cells have 46 and gametes have 23. For simplicity, the dashed arrows indicate the lineages of cells that we do not follow.

By contrast, spermatogenesis occurs continually throughout a male's lifetime upon reaching sexual maturity (i.e. puberty). At a bare minimum, a developed sperm cell has undergone 34 rounds of mitosis (30 leading to the formation of the stem cell and 4 after the stem cell) and 1 round of meiosis. But there are also additional rounds of mitosis to take into account as the result of the stem cells continually dividing to maintain the sperm supply. With these stem cells dividing every 16 days after puberty, the number of genome duplications to make a man's sperm is dependent on the age of the man.

(c) How many genome replications have occurred to make a "typical" man's sperm? In this context, we consider that a "typical" male hits puberty at 15 and reproduces at 30 years old.

(d) Given your answer in **2c**, how many mutations do you expect this "typical" man to pass on to his children?

We have now estimated the total number of mutations that we expect the mother and the father to contribute, allowing us to determine the total number of mutations per human offspring.

(e) What is the total number of mutations we expect to accumulate in a human offspring? What are the relative effects of the mother and the father in this estimate?

(f) Make a plot of the number of mutations accumulated in the gametes as function of age for males and females. Make sure to graph the number of mutations in the egg and the

sperm on the same plot to better compare their relative effects.

4. Open reading frames in random DNA

Do problem 4.7 of PBoC shown in Figure 2. Note that Figure 1.4 from PBoC can be found in Figure 3 of this homework. Finally, note that (b) and (c) in the problem are asking you to compute the probability of having an ORF of *at least* N codons in length.

> **4.7 Open reading frames in random DNA**
>
> In this problem, we will compute the probabilities of finding specific DNA sequences in a perfectly random genome, for which we assume that the four different nucleotides appear randomly and with equal probability.
>
> **(a)** From the genetic code shown in Figure 1.4, compute the probability that a randomly chosen sequence of three nucleotides will correspond to a stop codon. Similarly, what is the probability of a randomly chosen sequence corresponding to a start codon?
>
> **(b)** A reading frame refers to one of three possible ways that a sequence of DNA can be divided into consecutive triplets of nucleotides. An open reading frame (ORF) is a reading frame that contains a start codon and does not contain a stop codon for at least some minimal length of codons. Derive a formula for the probability of an ORF having a length of N codons (not including the stop codon).
>
> **(c)** The genome of *E. coli* is approximately 5×10^6 bp long and is circular. Again assuming a that the genome is a random configuration of base pairs, how many ORFs of length 1000 bp (a typical protein size) would be expected by chance? Note that there are six possible reading frames.
>
> (Problem courtesy of Sharad Ramanathan)

Figure 2: Problem 4.7 from PBoC.

5. Dynamics of the constitutive promoter

In class, we determined that the rate of mRNA decay γ, and not the production rate r, dictates the time it takes for the mean mRNA number to reach its steady state value. Here, we further explore this conclusion that could be at odds with our initial expectations about the dynamics of the constitutive promoter.

(a) If r does not dictate the time to reach steady state, what aspect of the promoter dynamics does it determine? Solve for the mRNA concentration as a function of time for two different values of r such as 10 mRNA/min and 20 mRNA/min using an initial condition $m(t = 0) = 0$. Use $\gamma = 1$ /min. Plot mRNA number vs. time and show that r controls the

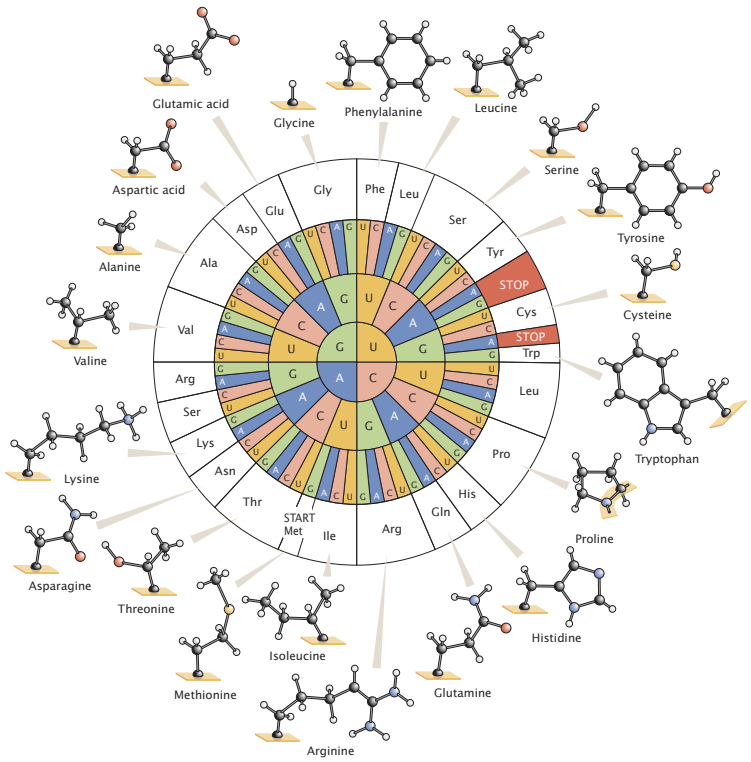

Figure 3: Genetic code. In this schematic representation, the first nucleotide in a coding triplet is shown at the center of the ring, the second nucleotide in the middle colored ring and the third nucleotide in the outer colored ring. In this representation of the genetic code, the four bases are adenine (A), cytosine (C), guanine (G) and uracil (U). Uracil is structurally very similar to thymine (T), and is used instead of thymine in messenger RNA. The amino acids corresponding to each group of triplets are illustrated with their names (outer ring) and atomic structures. Two amino acids, tryptophan and methionine, are encoded by only a single triplet, whereas others including serine, leucine, and arginine are encoded by up to six. Three codons do not code for any amino acid and are recognized as stop signals. The unique codon for methionine, AUG, is typically used to initiate protein synthesis.

initial slope.

(b) If r determines this initial slope, how come both curves take the same time to reach their steady state value? Plot the phase diagrams corresponding to both choices of model parameters and show that, while $r = 20$ mRNA/min has a faster initial increase in mRNA number, its steady state value is also larger such that the time it takes to reach steady state remains unaltered.