PHYS 176/276: Quantitative Molecular Biology
Problem Set #1
Due: **Friday, January 15, 2021**

1. **Making of the cell:** An *E. coli* cell is placed in a tube containing 5 ml sterile growth medium, consisting of 0.5% w/v glucose ($C_6H_{12}O_6$), 5mM ammonium (NH_4Cl), various inorganic ions, and kept shaken (for good aeration) in a 37° C water bath. After a short transient (to be neglected here), it is observed to grow exponentially with 60 min doubling time until it runs out of either the carbon or nitrogen source. Below you are asked to make a number of estimates using only the information provided here (you might check molar masses). Please report your numerical answers as well as the mathematical expressions.

   (a) Given that the dry weight of an *E. coli* cell is ∼0.3 pg, and that its chemical composition is ∼50% carbon and ∼14% nitrogen, estimate how many molecules of glucose and ammonium are required to make a new cell. What is the maximum cell density this tube of culture can reach? Express your answer in units of 'OD' (optical density, defined as light absorption through 1 cm of culture at a wavelength of 600nm), with 1 OD$\sim 10^9$ cells/ml. How long would it take for the tube to reach this density starting from a single cell? from 10^7 cells/ml (the typical starting culture density)? When the culture is saturated, how much volume does each cell have to itself (typical size of an *E. coli* cell with given growth rate is $1\mu m$)? How does this compare to the volume of an *E. coli* cell?

   (b) Let us next estimate the energy cost of making the cell. The dominant cost of biosynthesis turns out to be in protein synthesis. It takes 4 ATP molecules to extend a nascent polypeptide chain by one amino acid. Given that about half of the cells dry weight is in proteins, how many ATP does it cost to synthesize all the proteins in a cell? (You may assume that half of the protein mass is in Carbon and that one amino acid contains on average 5 carbon atoms.) Given the ATP hydrolysis energy of ∼30 kJ/mol, what is the energy cost (in Joules) to synthesize all the proteins in one cell? In all cells in the tube?

   (c) The ATP concentration is maintained at ∼4mM in the cell. From your answer to part (b), estimate the rate ATP is being drained to perform protein synthesis. How long would it take to deplete the ATP pool if it is not being supplied? Theoretically, one glucose molecule can maximally generate ∼30 ATPs using respiration under aerobic conditions. Estimate the minimal number of glucose molecules needed to supply the energy consumption for protein synthesis. Compare this amount to the need of glucose as building blocks of cell.

   (d) It turns out that *E. coli* actually generates energy from glucose very inefficiently even in aerobic conditions, largely fermenting glucose (generating 12 ATP and excreting 2 acetate molecules per glucose) rather than burning up all the carbon into CO_2 as it would occur in respiration. Assuming that glucose used to generate energy is completely via fermentation (instead of respiration), estimate again the minimal number of glucose molecules needed to supply the energy requirement for biosynthesis, and compare this amount to the need of glucose as building blocks of cell. What is the total glucose consumption rate in this case? And what is the rate of acetate accumulation in the medium? Find the acetate concentration in the medium when growth stops. Look up the chemical properties of acetate and express the answer in volume fraction. What is the pH of the medium at that point? Compare to table vinegar which tends is $4 \sim 8\%$ of acetate by volume.

2. **Scales:** In this problem, we will try to get a handle on the magnitude of various sizes and durations for molecules and processes in the cell. Please report your numerical answers as well as the mathematical expressions.

   (a) How many pages of a typical novel would it take to spell out the entire genome of *E. coli* ? of the human genome? Assuming the information content of the genome is comprised mostly in the genes, how many pages of a typical novel would the information be comparable to? (You may assume each gene encodes a protein 300 amino acid in length, and the alphabet size of 20 amino acids for the proteins is comparable to the 26-letter of the English alphabet. Make some reasonable assumptions about page content.)

   (b) Suppose a single *E. coli* cell invades into the intestine of a newborn infant and replicates at the rate of one doubling every 60 min. How long would it take for the *E. coli* colony to reach the size of 10^{14} cells (the estimated steady-state abundance of bacteria in the gut)? How much would these *E. coli* cells weigh? If the colony continues to grow at this rate, how much more time would it take for the colony to reach the mass of a 8 lb infant? How much more time would it take to reach the mass of the earth?

   Scale an *E. coli* cell to the size of a football stadium, say 100m by 100m by 200m and find the corresponding sizes of the following objects. Try to make analogies with objects with familiar sizes.

   (c) What would be the diameter and length of genomic DNA (4.6×10^6 base pairs) after scaling? what about distance separating each base pair, or alternatively the size of a base pair? If you wrap the DNA like racetrack within the stadium, how many laps would it take? How long would it take for you to bike half the length of this chromosome? (It takes DNA polymerase 40 minutes.)

   (d) If you cut the chromosome into 200m pieces and hang these pieces lengthwise across the stadium in a square array, how many pieces will there be and what would be the distance between the nearby pieces?

   (e) What would be the size of a typical small protein after scaling? the sizes of small molecules (e.g., glucose, NH_3) that the proteins work on? the sizes of ribosome and RNA polymerase (RNAp)?

(f) Compare the crowdedness inside a cell (with $\sim 3 \times 10^6$ proteins per cell) with the typical capacity of a stadium. (A large stadium seats 100,000 people.) If the proteins fill the stadium uniformly in 3d, what would be the inter-protein spacing?

Next, scale the cellular doubling time of 60 min to the time of 1 year to construct a stadium. Find the corresponding time scales for the following processes.

(g) Proteins in a cell diffuse roughly as $l \approx \sqrt{Dt}$ with $D \approx 10\mu m^2/s$. How do you have to scale the diffusion rate to have a comparable effect of diffusion within the stadium? How long would it take for a small protein to diffuse across the stadium?

(h) What would be the speed of DNA replication after scaling? How long does DNAp spend on each nucleotide?

(i) How long would it take for the RNAp to transcribe a typical gene (1000 nt long) if the real speed of RNAp is 48 nt/s ? and for the ribosome to translate it into protein (real speed is 16aa/s)?

3. **Transcription and translation rates:** In parts (a)-(f), deduce the transcription and translation rates for typical (non-ribosomal) genes in the exponential growth phase. Take the doubling time to be 60 min, and use the average copy number of a gene to be 2. Please report your numerical answers as well as the mathematical expressions.

Table 1: Pertinent values

ε_{tsx}	maximal transcription speed	~48 nt/sec
ε_{tsl}	maximal translation speed	~16 aa/sec
ℓ_{RNAp}	physical size of RNAp	~55 nt
ℓ_{Rb}	physical size of ribosome	~35 nt
τ_{mRNA}	half life of mRNA	~2 min
δ_{mRNA}	mRNA degradation rate	$\ln 2/\tau_{\text{mRNA}}$
T	cell doubling time	60 min
μ	dilution rate due to growth	$\sim\ln 2/T$
g	average gene copy number	2
$G_{E.coli}$	total genes in *E. coli*	4500
l_{gene}	average gene length	300 aa

(a) Given the maximal speed of transcriptional elongation by RNAp (~48 nt/sec) and the physical size of RNAp (covers ~55 nt), find the maximal rate at which full-length mRNA transcripts can be synthesized. [Hint: think about how frequently can a new transcript be initiated.]

(b) Given the half-life of 2 min for a typical transcript, what is the maximal copy number for each type of such transcripts in the steady state (of balanced exponential growth)? If every gene is transcribed at this rate, how many RNAp would be needed?

(c) One actually finds only 2-3 transcripts for a typical (non-ribosomal) gene in an *E. coli* cell. How frequently are these genes being transcribed? Suppose half of the genes in the genomes are transcribed at this rate (most other genes are not actively transcribed), how many RNAp would be needed to do the job?

(d) For a given mRNA transcript, what is the maximum number of proteins that can be synthesized from it within its lifetime? (Note: the maximal elongation speed of a ribosome is $\sim$16 amino acids/sec. Its size is $\sim$35 nt.)

(e) Given the steady state concentration of 2-3 transcripts for each of the actively transcribed genes in the cell, what would be the steady state concentration of the (non-ribosomal) proteins, assuming a maximum translation speed and that most of these proteins are not actively degraded, but are instead only diluted due to cell growth and division? How many ribosomes would it take to maintain such a concentration?

(f) The actual cytoplasmic protein concentration is only about 3×10^6/cell, corresponding roughly to 1000$\sim$1500 per actively transcribed gene. How frequently does this suggest that each mRNA transcript is being translated? Typically, how many ribosomes would you find on a transcript? How many total ribosomes are needed to maintain the observed protein concentration in the steady state?

Now, consider ribosomal RNA (rRNA) synthesis. The transcriptional elongation speed of RNAp on rRNA genes is $\sim$85 nt/sec.

(g) Given there is an average of 20 copies of rRNA genes at 60 min doubling time, what is the maximal number of rRNA the cell can synthesize in one doubling? How many RNAp would it take (length of a typical rRNA gene is 4500 nt)?

(h) There are actually 15,000 ribosomes per cell at 60 min growth. What is the initiation rate of rRNA synthesis at each rRNA gene? what is the average separation between successive RNAp? and how many RNAp are involved in rRNA synthesis?