CHEMENG 355 Spring 2018

Problem Set 1 Due April 19, 2018

Question 1 – Rates of transcription and translation

In class we discussed the importance of manufacturing the human peptide hormone insulin using a recombinant host. One way to make insulin in *Escherichia coli* mentioned in class is to produce the two subunits separately and then combine them *in vitro* (*i.e.* outside the cell). Another method involves production of insulin in *E. coli* as a single peptide and then digestion of proinsulin after it is isolated to yield the final form of the peptide hormone with the two subunits linked by disulfide bonds. The heterologous production of proteins in *E. coli* commonly features a promoter borrowed from the T7 bacteriophage (a virus that infects *E. coli*), which when turned on can cause a protein of interest to account for up to half of all the mRNA transcripts and translation products in a cell.

Let's examine insulin production in an *E. coli* batch fermentation (using the method where insulin is produced as a single peptide) and assume the culture is saturated and cellular growth rate is negligible (OD_{600} = 10, 5 liter culture volume; OD_{600} of 1.0 for *E. coli* corresponds to 5 x 10^8 cells per ml). Further estimate that ATP "usefulness" is only 50% (BNID 102605) - that is, half of all ATP made in the cell is lost in membrane potential buildup and leakage - and that this ATP is made by 3000 ATP synthases per cell (von Meyenburg, 1984).

A. Using estimates for transcription and translation energetics in bacteria discussed in lecture, calculate the minimum time required to produce a day's worth of insulin for a single patient (68 units of insulin per day, where 1 unit of insulin is defined as 34.7 ug of pure crystalline insulin) if ATP availability limits the total amount of insulin that can be made. Assume that each ATP synthase can produce 100 ATP/second.

B. From a biotechnological standpoint, does this calculated time make sense? Do you think it is reasonable to assume ATP availability alone limits the rate of protein synthesis? Why?

Let us instead analyze the production on a transcriptionally-limited basis. Assume that transcription of insulin is induced at stationary phase (such that every cell in the culture is making protein); mRNA transcripts have an average lifetime of 8 minutes; and each recombinant *E. coli* cell contains 60 copies of the insulin-gene-harboring expression plasmid. Further assume that translation and transcription initiation rates are maximal: Translation occurs at 48 nucleotides per second, with 120 nucleotides between each ribosome on the mRNA on average. Transcription occurs at 50 nucleotides per gene per second.

C. Keeping in mind the original culture parameters above, how many transcripts per cell are required to synthesize a day's worth of insulin for a single patient? How does this compare to the average number of mRNA transcripts per gene in the *E. coli* genome?

D. In the transcriptionally-limited regime, what is the length of time required for the culture to produce a day's worth of insulin for a single patient? Is insulin production likely limited by ATP availability or transcription?

Part A (7 points)
How many molecules of insulin are needed?

Use the molecular mass of insulin here because the final dose insulin, not proinsulin:

$$\frac{68\ Units}{1} * \frac{34.7\ ug}{Unit} * \frac{g}{10^6 ug} * \frac{1\ mol}{5808\ g} * \frac{6.02 * 10^{23}\ proteins}{mol} = 2.45 * 10^{17} proteins$$

How many peptide bonds are required for each dose of insulin? How many molecules of mRNA are needed?

$$\frac{2.45 * 10^{17}\ proteins}{1} * \frac{86\ peptide\ bonds}{protein} = 2.10 * 10^{19}\ peptide\ bonds$$

How many proteins can each mRNA make? Since the spacing of the ribosome on a transcript is one ribosome every 120nt, and 48nt are translated every second, we know that translation is initiated every 2.5 seconds. Since the message exists for 8 minutes, with initiation every 2.5 seconds, and an mRNA lifetime of 8 minutes, 192 proteins are made for every transcript.

$$\frac{2.45 * 10^{17}\ proteins}{1} * \frac{transcript}{192\ proteins} = 1.27 * 10^{15}\ transcripts$$

OK if used 3 minute lifetime for mRNA instead (number given in class, not specified in the problem until part c, not too consequential for the rest of the math):

$$\frac{2.45 * 10^{17}\ proteins}{1} * \frac{transcript}{72\ proteins} = 3.40 * 10^{15}\ transcripts$$

How much ATP is required for transcription and translation?

$$\left(\frac{2.10 * 10^{19}\ peptide\ bonds}{1} * \frac{4.2\ ATP}{peptide\ bond}\right) + \left(\frac{1.27 * 10^{15}\ transcripts}{1} * \frac{86\ codons}{transcript} * \frac{6\ ATP}{codon}\right)$$
$$= 8.89 * 10^{19}\ ATP$$

How many cells are there? How many ATP synthases are there? What is the overall ATP Synthesis rate?

$$\mathbf{\frac{5 * 10^{8}\ cells}{mL * OD_{600}} * \frac{5000\ mL}{1} * \frac{10\ OD_{600}}{1} = 2.5 * 10^{13}\ cells}$$

$$\mathbf{2.5 * 10^{13}\ cells * \frac{3000\ ATP\ Synthases}{cell} = 7.5 * 10^{16}\ ATP\ synthases}$$

$$\mathbf{7.5 * 10^{16} ATP\ Synthases * \frac{100\ ATP}{s * ATP\ Synthase} = 7.5 * 10^{18}\ ATP/second}$$

In the last step, we divide the ATP synthesis rate by 2 because half of the ATP is going towards our protein of interest, while the other half is going towards translating other transcripts in the cell (the problem statement indicated that half the transcipts will be our protein of interest.

We divide the ATP synthesis rate by 2 again because the problem statement says half of the ATP is lost to membrane potential buildup and leakage.

$$\frac{8.89 * 10^{19}\ \ ATP}{1} * \frac{second}{7.5 * 10^{18}\ ATP} * \frac{4\ ATP}{useful\ ATP} = \mathbf{47.4\ seconds}$$

Part B (2 points)

This is unreasonable from a biotechnological perspective – we wouldn't have any issues ever again making insulin. A 5 L fermenter is very small from an industrial perspective, so producing one dose so quickly is not a reasonable expectation. ATP availability is probably not rate-limiting.

Part C (2 points)

We found the total number of proteins and the number of cells and the number of proteins per transcript in part A of this problem:

$$\frac{2.45 * 10^{17}\ proteins}{1} * \frac{transcript}{192\ proteins} * \frac{1}{2.5 * 10^{13}\ cells} = \mathbf{51\ transcripts/cell}$$

This is much higher than the average number of transcripts that exist per gene in the E. coli genome at any given time. The average number of transcripts per gene is ~0.5 transcripts/gene. This shows that this is a very strong promoter (1pt)

Part D (5 points)

How many transcripts are produced for every gene in order to make the full dose of insulin?

$$\frac{51\ transcripts}{cell} * \frac{cell}{60\ genes} = \frac{0.851\ transcripts}{gene}$$

Insulin → 86 aa protein = 258 nt/transcript (1pt)

Transcription rate is 50 nt/s:

$$\frac{0.851\ transcripts}{gene} * \frac{256\ nucleotides}{transcript} * \frac{gene * second}{50\ nucleotides} = \mathbf{4.4\ seconds\ (2\ points)}$$

This rate is once again very fast. If we have to choose between the ATP limited regime and the transcriptionally limited regime, the ATP limited regime makes more sense. However, the extremely fast times indicate that in reality, some other factor is likely limiting. Possible candidates for the true limiting factor include mass transport (oxygen to electron transport chain, carbon source into cells, nitrogen source into cells) and building block availability (all of the amino acids incorporated into the protein and the nucleotides making up the mRNA need to be synthesized from the carbon and nitrogen sources available before being incorporated). Producing this amount of a single protein would be extremely taxing on all aspects of cell growth, so maximal ATP synthesis or transcription and translation rates are likely bad assumptions to make in real life. (2 points)

Question 2 – Engineering translation

In recent years, several research groups have undertaken efforts to engineer *E. coli* and yeast to produce proteins that contain unnatural amino acids, reasoning that these additional amino acids could endow proteins with new and useful functions (e.g., in catalysis and imaging). The most commonly used strategy is to 'expand the genetic code' by using one of the three stop codons to encode the unnatural amino acid.

A) What are the two biomolecules that need to be engineered for such an effort? *Hint: one is an enzyme and the other is a nucleic acid.*

Answer: A new tRNA and an aminoacyl-tRNA synthetase that charges this tRNA with the unnatural amino acid. (2 points)

B) What are two key challenges of engineering the enzyme from part A starting from a native variant of the same enzyme? One should address its ability to function as desired, and the

other should address the potential problems of introducing a new enzyme of this sort into a cell.

1) Need to get the aminoacyl-tRNA synthetase to recognize a new substrate (2 points)

2) Need to ensure the engineered aminoacyl-tRNA synthetase does not mis-charge other native tRNAs (2 points)

C) Bonus: Add a third key challenge that is a more general concern about the potential effects on the translation of other cellular proteins.

Some subset of other cellular proteins will no longer be translated properly (due to a competing engineered tRNA for the native stop codon). For further reading, please see *Science*, 2001, 498-500. (1 points)

Question 3 – Making and using ATP

One of the most fascinating aspects of using cells as biocatalysts is that they can self-replicate. For example, ATP synthase produces ATP, which in turn is required for its own production. However, cellular metabolism of growth and protein production comes at an energetic cost that competes with product formation in an engineered strain.

a) Using data provided in the Weissman paper (Cell 157, 624–635, April 24, 2014), estimate how long it takes a cell to produce 1000 new ATP synthases. Assume that a generation time for an *E. coli* cell is 20 minutes.

b) How long does an ATP synthase need to operate to generate enough ATP to pay for the cost of the protein translation that generated it? (Assume an ATP synthase turnover rate of 100 ATP/second)

c) In E. coli, ATP Synthase is composed of several subunits (AtpA-AtpH). The subunits are all produced from a single mRNA transcript but are not needed in a 1:1 ratio. From Figure 2 of the paper, which subunit is the most abundant in ATP synthase? Are the subunits synthesized in a 1:1 ratio, or does protein synthesis more accurately reflect the stoichiometry of the subunits in the ATP synthase complex? What mechanism is likely to account for any differences in synthesis rate?

Solution:

a) (3 points) AtpE is synthesized at a rate of $1.1\text{x}10^4$ molecules/generation from Fig. 2A, right panel. Since the key result of the paper is that the production rate scales linearly with subunit stoichiometry, the time to synthesize a proportional number of each subunit will be about the same, and since this can happen in parallel for the subunits, we only need to calculate the time to synthesize 1000 full sets of any given subunit and we will be approximately correct.

$$1000\ \text{ATP synthases} * \frac{10\ \text{AtpE subunits}}{\text{ATP synthase}} * \frac{\text{generation}}{1.1\text{x}10^4\ \text{AtpE subunits}} * \frac{20\ \text{min}}{\text{generation}} = 18.2\ \text{min}$$

Other subunits can also be used (and estimates for the number of subunits/generation will vary), but again, answers should all be close to the 18 minute estimate because the number of subunits per generation is proportional to the number of subunits in an ATP synthase complex.

b) (3 points) We are told to assume many of the quantities needed to answer this question. To determine the total cost of protein synthesis to make an ATP synthase, we can make an estimate based on the size of the total protein (520 kDa) and the estimate for average molecular weight of an amino acid (*AA*), given as 110 g/mol.

$$520{,}000\ \text{Da} * \frac{AA}{110\ \text{Da}} * \frac{4.2\ \text{ATP}}{AA} * \frac{\text{s}}{100\ \text{ATP}} = 199\ \text{s}$$

Also ok to use the number of amino acids in the ATP synthase complex, but for a protein as large as the ATP synthase complex, the 110 Da/amino acid estimate is probably very good. It is more important to calculate the average for a specific protein when the protein is small. (using 4 as the ATP/AA is a ½ point deduction)

c) (2 points) AtpE is the most abundant (10x more abundant than AtpB/C/G/H). The subunits are not synthesized in a 1:1 ratio; they are synthesized in a ratio more proportional to their stoichiometry in the ATP synthase complex (as indicated in the chart on the right side of Figure 2A). The paper does not indicate the actual reason for differences in synthesis rate, but anything reasonable can be accepted. mRNA amount, however, is NOT an acceptable answer since the blue chart in Figure 2A indicates that the amount of mRNA is relatively constant for the entire mRNA fragment. Possible answers include RBS strength, mRNA secondary structure, and translational efficiency increases or drops due to codon usage.