

Problem Set 4 on Central Dogma

(due Monday Oct. 19th, in class or by 5:00pm in Prof. Goutte's office)

1. Define the following terms as they are used in regular English language AND give the special meaning each term has when applied to the living cell.
 - a. Replication: **The action or process of reproducing or duplicating. In cells, the process of duplicating cellular material, such as DNA, to make two identical copies to pass on to daughter cells.**
 - b.
 - c. transcription: **a: an arrangement of a musical composition for some instrument or voice other than the original b : a recording (as on magnetic tape) made especially for use in radio broadcasting. In cells, the process of constructing a messenger RNA molecule using a DNA molecule as a template with resulting transfer of genetic information to the messenger RNA.**
 - d. translation: **a : a rendering from one language into another. In cells, the process of forming a protein molecule at a ribosomal site of protein synthesis from information contained in messenger RNA.**
 - e. splicing: **a : to unite (as two ropes) by interweaving the strands b : to unite (as lengths of magnetic tape) by overlapping and securing together two ends. In cells, the cellular process by which intervening sequences (introns) are removed from mRNA transcripts, and remaining exons are rejoined together.**
2. Now for some numbers...
 - a. How many different types of nucleotides are found in RNA? **3**
 - b. How many different types of amino acids are found in proteins? **20**
 - c. How many nucleotides are needed to code for a single amino acid? **3**
 - d. How many different codons are there in the genetic code? **64**
 - e. How many codons code for amino acids? **61**
3. Myoglobin is 153 amino acids long.
 - a. What would be the minimum number of nucleotides required to encode myoglobin?
 $(153 \times 3) + 3 = 462$
 - b. Why might the size of the myoglobin primary RNA transcript be larger than your answer above? **Because the primary transcript has large introns which are removed before the mRNA is used for translation of myoglobin.**

4. Use the genetic code to identify which of the following nucleotide sequences would code for the polypeptide sequence arginine-glycine-aspartate:
 - a. 5'-AGAGGAGAU-3'
 - b. 5'-ACACCCACU-3'
 - c. 5'-GGGAAAUUU-3'
 - d. 5'-CGGGGUGAC-3'

Sequences **a** and **d** both code for the peptide Arg-Gly-Asp. Because the genetic code is redundant, different nucleotide sequences can encode the same amino acid sequence.

5. Explain why if you are given a protein sequence, you cannot predict the exact RNA sequence that was used by the cell to generate that sequence.

Because there is more than one codon that can specify most amino acids, therefore you never know which one was used for a particular protein, unless you can look at the DNA or RNA sequence.

6. Consider the hypothetical protein represented by single letter amino acid designations:
A-M-H-E-R-S-T.
 - a. Propose one RNA sequence that *could* encode this sequence (don't worry about start codons for this exercise).

Lots of Possibilities here!

- b. Can you figure out mathematically how many different RNA sequences could code for this mini protein?

I've indicated the number of possible codons for each amino acid:

A(2), M(1), H(2), E(2), R(6), S(6), T(4).

In order to figure out the number of possible different mRNAs, you simply multiply these numbers (2x1x2x2x6x6x4), and you get 1152 possible RNA sequences that would code for "AMHERST"... now, just when you are feeling pretty good about your odds of having found a correct RNA sequence to encode AMHERST, you should consider the following: how many different RNA sequences are possible for this length (21 nucleotide-long) RNA?? (to figure this out: you've got 4 possible nucleotides at each of 21 spots, so that's 4^{21}) Now you'll see that if you guessed randomly you'd have less than one in a billion chance of getting it right!

7. Here is the beginning of a protein-encoding gene sequence.

5' -ATGAAGTTTGGCACTTAA-3'

3' -TACTTCAAACCGTGAATT-5'

 - a. Give the RNA transcript that would be transcribed off of the bottom strand:

5' -AUGAAGUUUGGCACUUA-3'
 - b. Translate this RNA sequence into a protein sequence.

MET-LYS-PHE-GLY-THR
 - c. Give the RNA transcript that would be transcribed off of the top strand:

5' -UUAAGUGCCAAACUUC-3'
 - d. Translate this RNA sequence into a protein sequence.

This one does not have a starting MET, so technically it would not get translated; if it did, the sequence would have been: LEU-SER-ALA-LYS-LEU-HIS

7 continued

- e. Let's assume that the bottom strand is the strand that is used as a template strand when this gene gets transcribed. What would be the effect on the final protein product if a mutation caused the following single base-pair insertion:

5' -ATGAAG**A**TTTGGCACTTAA-3'
3' -TACTT**C**TAAACCGTGAATT-5'

because the reading frame of the protein is changed, all codons after the lysine are different, hence you get a completely different protein: MET-LYS-TRP-HIS-LEU

- f. What would be the effect on the final protein product if a mutation caused the following single base-pair substitution:

5' -ATGAAGTT**C**GGCACTTAA-3'
3' -TACTT**C**AAAGCCGTGAATT-5'

no change: MET-LYS-PHE-GLY-THR

- g. What would be the effect on the final protein product if a mutation caused the following single base-pair substitution:

5' -ATGAAGTTT**C**GGCACTTAA-3'
3' -TACTT**C**AAAGCCGTGAATT-5'

a protein with one amino acid change: an arginine replaces a glycine; because a tiny a.a. is being replaced by a large charged a.a. this change is likely to affect the overall 3-D structure and hence the stability/function of the protein: MET-LYS-PHE-ARG-THR

- h. What would be the effect on the final protein product if a mutation caused the following single base-pair substitution:

5' -ATG**T**AGTTTGGCACTTAA-3'
3' -TAC**A**TCAAACCGTGAATT-5'

this mutation changes the second codon into a stop codon which has the effect of terminating translation, thus no protein would be synthesized on this message (unless there happens to be another AUG codon available farther along, and in this case the outcome depends on how far away it is, and whether or not it is in the same reading frame as the original start codon).

- i. Which of the above mutation(s) would you expect to be the most severe in terms of the overall effect on the person carrying such a mutation? Explain.

(e) and (h) would be the most severe. There is a slight chance that in (h) another in-frame start codon exists, and thus the original protein might have a chance of being made with some missing a.a. at the beginnings, but this is not the most likely scenario. (g) is only one a.a. different from the original, however if this happened to be critical for protein folding, it is possible that this mutation would be just as bad as (e) or (h). (f) would have absolutely no effect on the person carrying this mutation.