

Name: _____

ID : _____

ECS 129: Structural Bioinformatics
Midterm-makeup
February 10, 2022

Notes:

- 1) The midterm is open book, open notes.
- 2) You have 45 minutes, no more: I will strictly enforce this.
- 3) The midterm is divided into 3 parts and graded over 90 points.
- 4) You can answer directly on these sheets (preferred), or on loose paper.
- 5) Please write your name at the top right of each page you turn in!
- 6) Please, check your work! **Show your work** when multiple steps are involved.

Part I (6 questions, each 10 points; total 60 points)

(These questions are multiple choices; in each case, find the **most plausible** answer)

- 1) A protein sequence contains one ASP residue. You want to create a new protein sequence, with this ASP being replaced with a TYR. To do this, you first generate the cDNA corresponding to the original protein (with your own choice for the codons you use), then mutate this cDNA to get the sequence corresponding to the new protein. What is the minimum number of mutations needed?
 - a) 0
 - b) 1**
 - c) 2
 - d) 3
 - e) None of the above

ASP can be represented with the codon GAU; mutating G with U, you get the codon UAU which codes for TYR

- 2) You have sequenced a small peptide, but unfortunately the results are only partial: Met Trp XXX Trp Met, where XXX could be any of the twenty amino acids. You are told, however, that the DNA strand that codes for this peptide contains exactly 40% of thymine (not considering the STOP codon). Can you find which amino acid in this peptide could be XXX?
 - a) Phe**
 - b) Pro
 - c) Lys
 - d) Ala

The small peptide contains 2 Met that are each coded by AUG, 2 Trp that are encoded by UGG, and one unknown amino acid. The DNA strand encoding this peptide is of the form

ATG TGG XYZ TGG ATG (1)

As there are exactly 40% of thymine over those 15 nucleotides, there is a total of 6 thymine, and therefore 2 thymine among XYZ. Of the four options, only Phe has such a codon, in fact this codon is TTC.

Name: _____

ID : _____

3) Only one of these DNA sequences can form a strand that can pair with itself to form a double stranded molecule. Which one?

- a) 5'-GCGCAAGCGC-3'
- b) 5'-GCGCATGCCC-3'
- c) 5'-GCGCATGCGC-3'
- d) 5'-GGGGATTACCCC-3'
- e) 5'-ATGCATGCATGC-3'

c) is the only one that is its own complementary.

4) How many DNA coding sequences (including the START and STOP codons, but no introns) could lead to the following protein sequence:

Met- Ile-Ser-Trp-Leu-Gln-Trp-Ala, assuming the standard genetic code?

- a) 2592
- b) 864
- c) 1728
- d) 8
- e) Not enough information

Count number of codons for each amino acid (3 for stop codons):

1 (Met) x 3 (Ile) x 6 (Ser) x 1 (Trp) x 6 (Leu) x 2 (Gln) x 1 (Trp) x 4 (Ala) x 3 (STOP) = 2592

5) Assume that the human genome contains 6 billion bases (i.e. 6×10^9) and that each base weighs on average 100 Da, where 1 Dalton = 1.66×10^{-27} kg. What would be the weight of a DNA molecule that would store all the collections in the Library of congress, assuming that those collections contains 3 “peta characters” (i.e. 3×10^{15} characters) and that each letter needs to be stored over 6 bases?

- a) 4.98 grams
- b) 29.8 grams
- c) 2.98×10^{-6} grams
- d) 4.98×10^{-7} grams
- e) 2.98×10^3 grams

Mass = 3×10^{15} (characters) x 6 (base / character) x 100 x 1.66×10^{-27} (kg) x 1000 (g/kg) = 2.98×10^{-6} g

6) You are given a single strand S1 of DNA. You are told that: (i) if you mutate one Adenine of S1 to a Thymine, the mutated DNA contains as many purine as pyrimidine, and (ii), if you mutate two Thymine of S1 into Adenine, you then have twice as many purines as pyrimidines. What is the length of S1?

- a) 18
- b) 20
- c) 21
- d) 54
- e) Not enough information

Name: _____

ID : _____

Let A, T, G, and C be the number of Adenine, Thymine, Guanine, and Cytosine in S1. We know that:

$$(A-1+G) = T+1+C$$

$$A+2+G = 2(T-2+C)$$

Let us define $A+G = X$ and $T+C = Y$. Then

$$X-1 = Y+1$$

$$X+2 = 2(Y-2)$$

$$\text{Hence } X = Y+2$$

$$\text{And } Y+4 = 2Y - 4$$

$Y = 8$ and $X = 10$; therefore the length of S1 is $X+Y = 18$.

Name: _____

ID : _____

Part II (4 questions, each 5 points: total 20)

Below is the double-stranded DNA sequence of part of a hypothetical yeast genome, which happens to contain a very small gene.

```
5' – TATAAAGAGCCATGCATGAACTGGATAAAAGGCTCTGAGAATTTATCTCTAG– 3'
      |||||
3' – ATATTTCTCGGTACGTACTTGACCTATTTTCCGAGACTCTTAAATAGAGATC– 5'
```

- a) Which strand of DNA shown, the top or the bottom, is the coding strand? Justify your answer

Only the top strand contains a START codon in phase with a STOP codon:

5' –ATG CAT GAA CTG GAT AAA AGG CTC TGA–3'

- b) What is the mRNA sequence corresponding to the ORF for the gene?

5' – AUG CAU GAA CUG GAU AAA AGG CUC UGA –3'

- c) What is the sequence of the protein produced from the mRNA in (b)? Label the N and C termini.

Nter- MHELDKRL-Cter

- d) Propose a single nucleotide mutation (substitution, insertion, or deletion) in the ORF corresponding to that protein that would lead to the new shorter protein sequence Met His.

Insertion of a T after CAT, leading to a new stop codon.

5' –ATG CAT TGA

Name: _____

ID : _____

Part III: (one question, 10 points)

We want to perform a simple “back of the envelope” calculation of the number of cells in a human body. We start first by estimating the average volume of one human cell. We use for that the simple, rough equation:

$$\text{Number of proteins per cell volume} \left\{ \frac{N}{V} = \frac{C_p}{I_{aa} \times m_{aa}} \right.$$

Average number of aa per protein Average mass of aa

With the following:

- $C_p = 0.2 \text{ g/ml}$
- $I_{aa} = 300$
- $m_{aa} = 100 \text{ Daltons}$
- $1 \text{ Dalton} = 1.66 \cdot 10^{-27} \text{ kg}$
- $1 \text{ ml} = 1 \text{ cm}^3 = 10^{12} \mu\text{m}^3$

Assuming that a human cell contains on average $5.0 \cdot 10^9$ proteins, what is its size (in μm^3)?

$$\text{Volume} = \frac{N \times I_{aa} \times m_{aa}}{c_p}$$

$$\text{Volume} = \frac{5 \cdot 10^9 \times 300 \times 100 [\text{Da}] \cdot 1.66 \cdot 10^{-24} \frac{[\text{g}]}{[\text{Da}]}}{0.2 \frac{[\text{g}]}{[\text{ml}]} \cdot 10^{-12} \frac{[\text{ml}]}{[\mu\text{m}^3]}}$$

$$\text{Volume} = \frac{5 \cdot 10^9 \times 300 \times 100 \times 1.66 \cdot 10^{-24} [\mu\text{m}^3]}{0.2 \times 10^{-12}}$$

$$\text{Volume} = \frac{5 \cdot 10^9 \times 300 \times 100 \times 1.66 \cdot 10^{-24} [\mu\text{m}^3]}{0.2 \times 10^{-12}}$$

$$\text{Volume} = 1245 \mu\text{m}^3$$

Now assume for simplicity that the mass density of a human cell is 1 g/cm^3 , and that an average human weighs 70 kgs, that all this weight is assigned to cells, how many cells are there in this average human?

$$N_{\text{cell}} = \text{mass} / (\text{density} \times \text{volume})$$

$$N_{\text{cell}} = \frac{70 \times 10^3 [\text{g}]}{1 \frac{[\text{g}]}{[\text{ml}]} \cdot 10^{-12} \frac{[\text{ml}]}{[\mu\text{m}^3]} \times 1245 [\mu\text{m}^3]}$$

$$N_{\text{cell}} = \frac{70 \times 10^3}{10^{-12} \times 1245} = 5.6 \cdot 10^{13}$$

Name: _____

ID : _____

Appendix:

Appendix A: Genetic Code

	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met/Start	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G