

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

- 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

- 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'

Name: _____

ID : _____

- 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
- 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
- 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 : _____

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' — TATAAAGAGCCATGCATGAACTGGATAAAAAGGCTCTGAGAATTTATCTCTAG— 3'
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3' — ATATTTCTCGGTACGTACTTGACCTATTTTCCGAGACTCTTAAATAGAGATC— 5'
```

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

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

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

- 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.

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:

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

Protein mass per volume ↓

↑ *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)?

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?

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