

Name: \_\_\_\_\_

ID : \_\_\_\_\_

**ECS 129: Structural Bioinformatics**  
**Midterm- Solutions**  
**February 2, 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 2 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 specific codon for Leu can be converted to either Ser, Val, or Met by a single nucleotide substitution. What is this codon?
  - a) UUA
  - b) UUG**
  - c) CUU
  - d) CUG
  - e) CUA

There are six codons for Leucine: UUA, UUG, CUU, CUC, CUA, and CUG.

Only UUA and UUG can be mutated to Ser with a single mutation, with the second U replaced with a C

Among those two (UUA and UUG), only UUG can be mutated to Met with a single mutation (first U replaced with A).

UUG can be mutated to Val by replacing the first U to G.

The solution is therefore UUG

- 2) You are told that the DNA strand that codes for the small peptide Met Trp Tyr Trp Met contains exactly 40% of thymine. Can you find which codon is used for the Tyr amino acid in this peptide?
  - a) UAU**
  - b) UAC
  - c) UGG
  - d) UAA
  - e) Not enough information

Name: \_\_\_\_\_  
ID : \_\_\_\_\_

The small peptide contains 2 Met that are each coded by AUG, 2 Trp that are encoded by UGG, and 1 Tyr that can be encoded by UAU or UAC. There are therefore two possible DNA strands encoding this peptide:

ATG TGG TAT TGG ATG (1)

or

ATG TGG TAC TGG ATG. (2)

Note that we do not consider the STOP codon, as it is not actually “coding”!

The first, (1), contains 6 T out of 15 bases, hence 40% of T, while the second one contains 5T out of 15 bases, hence 33% of T. The correct answer is therefore (1), i.e. Tyr is encoded by UAU.

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'-AAAAAAAAA-3'

b) 5'-AAAACCCC-3'

c) 5'-ATATATATA-3'

**d) 5'-GCATCCGGATGC-3'**

e) 5'-ATATATGGATATAT-3'

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

4) Which of the following statements is true?

a) DNA polymerase moves along the leading strand, reading it in the 3' to 5' direction

b) The new strand of DNA generated by the DNA Polymerase for both the lagging and the leading strands is synthesized in the 5' to 3' direction

c) The DNA polymerase needs to attach multiple times to the lagging strand, generating fragments that are subsequently attached together with a ligase

d) A DNA polymerase is bound to make mistakes; error rates depend on the type of DNA polymerase, and differ in different species

**e) All of the above**

f) None of the above

5) Assume that the human genome contains 6 billion bases (i.e.  $6 \cdot 10^9$ ) and that it weighs  $3.6 \cdot 10^{-12}$  grams. What would be the weight of a DNA molecule that would store the whole Encyclopedia Britannica, assuming that it contains 30 billion letters (i.e.  $3 \cdot 10^{10}$  letters) and that each letter needs to be stored over 3 bases?

a) 5.4 grams

b)  $5.4 \cdot 10^3$  grams

**c)  $5.4 \cdot 10^{-11}$  grams**

d)  $1.8 \cdot 10^{-11}$  grams

e)  $1.8 \cdot 10^3$  grams

Name: \_\_\_\_\_  
ID : \_\_\_\_\_

6 billions base pairs weigh  $3.6 \cdot 10^{-12}$  grams. We need 30 billion x3 base pairs to store the whole Encyclopedia Britannica, i.e 15 times more than the 6 billion base pairs of the human genome; this DNA molecule would weigh 15 times  $3.6 \cdot 10^{-12}$  grams, namely  $5.4 \cdot 10^{-11}$  grams

- 6) A single stranded DNA contains 15% Adenine, as many Guanines as Cytosines, and 40% of purines. What is the amount (in percent) of Thymine:
- a) 25%
  - b) 35%**
  - c) 40%
  - d) 15%
  - e) Not enough information

There are 15 % of Adenine, and 40% of purine (A+G): therefore there are 25% of Guanine. As there are as many Cytosine than Adenine, there are 25% of Cytosine. Therefore, the amount of Thymine is  $100-50-15 = 35\%$

Name: \_\_\_\_\_

ID : \_\_\_\_\_

**Part II (two questions, 10 points each: total 20)**

The following eukaryotic DNA sequence was given to you:

5'-GAGCC**ATG**CATTATCTAGATAGTAGGCTCT**TGA**GAAATTTATCT-3'.

- 1) You are told that this sequence codes for one gene. Find the longest “gene”, or open reading frame (ORF) corresponding to this DNA sequence. Transcribe this ORF into an RNA sequence and then into its corresponding protein sequence (label the 5' and 3' ends of the RNA, and Nter and Cter ends of the protein).

The complementary strand is:

5'- AGATAAATTCTCAGAGCCTACTATCTAGATAAGCATGGCTC-3'

There is therefore a single ATG in the strand and its complementary; this ATG is in phase with one TTGA (stop codon):

5'-ATG CAT TAT CTA GAT AGT AGG CTC TGA -3'

The corresponding RNA is:

5'-AUG CAU UAU CUA GAU AGU AGG CUC UGA -3'

And the corresponding protein sequence is:

Nter-Met-His-Tyr-Leu-Asp-Ser-Arg-Leu-Cter

- 2) We consider the same DNA as in questions 1) above. A mutation is found where a T base **is added immediately** after the T base shown in bold (with a larger font)

5'-GAGCCATGC**A**T**T**ATCTAGATAGTAGGCTCTGAGAATTTATCT-3'.

What would be the resulting RNA and protein sequences?

The mutated strand is:

5'-GAGCCATGCATTTATCTAGATAGTAGGCTCTGAGAATTTATCT-3'

This creates a new stop codon, TAG, in phase with the start codon, leading to the new RNA:

5'-AUG CAU UUA UCU AGA UAG-3'

With the protein sequence

Nter-Met-His-Leu-Ser-Arg-Cter

Name: \_\_\_\_\_

ID : \_\_\_\_\_

### Part III: (one question, 10 points)

We want to perform a simple “back of the envelope” calculation of the number of proteins within one cell of E-coli. 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:

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

Assuming that a bacterium has a volume of  $1 \mu\text{m}^3$ , how many proteins does this bacterium contain?

$$\text{Number of proteins} = \frac{0.2 \frac{[g]}{[ml]} \cdot \frac{1 [kg]}{1000 [g]} \cdot \frac{1 [ml]}{10^{-12} [\mu\text{m}^3]} [\mu\text{m}^3]}{300 \times 100 [Da] \cdot 1.66 \cdot 10^{-27} \frac{[kg]}{[Da]}}$$

$$\text{Number of proteins} = \frac{0.2}{3 \times 1.66} \times 10^8$$

$$\text{Number of proteins} = 4 \times 10^6$$

i.e. approximately 4 millions protein in one cell.

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