**ECS 129: Structural Bioinformatics**

**Midterm**

**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?
   1. UUA
   2. UUG
   3. CUU
   4. CUG
   5. CUA
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?
   1. UAU
   2. UAC
   3. UGG
   4. UAA
   5. Not enough information
3. Only one of these DNA sequences can form a strand that can pair with itself to form a double stranded molecule. Which one?
   * 1. 5’-AAAAAAAA-3’
     2. 5’-AAAACCCC-3’
     3. 5’-ATATATATA-3’
     4. 5’-GCATCCGGATGC-3’
     5. 5’-ATATATGGATATAT-3’
4. Which of the following statements is true?
5. DNA polymerase moves along the leading strand, reading it in the 3’ to 5’ direction
6. 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
7. The DNA polymerase needs to attach multiple times to the lagging strand, generating fragments that are subsequently attached together with a ligase
8. A DNA polymerase is bound to make mistakes; error rates depend on the type of DNA polymerase, and differ in different species
9. All of the above
10. None of the above
11. Assume that the human genome contains 6 billion bases (i.e. 6 109) and that it weighs 3.6 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 10 10 letters) and that each letter needs to be stored over 3 bases?
12. 5.4 grams
13. 5.4 103 grams
14. 5.4 10-11 grams
15. 1.8 10-11 grams
16. 1.8 103 grams
17. A single stranded DNA contains 15% Adenine, as many Guanines as Cytosines, and 40% of purines. What is the amount (in percent) of Thymine:
18. 25%
19. 35%
20. 40%
21. 15%
22. Not enough information

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

The following eukaryotic DNA sequence was given to you:

5’-GAGCCATGCATTATCTAGATAGTAGGCTCTGAGAATTTATCT-3’.

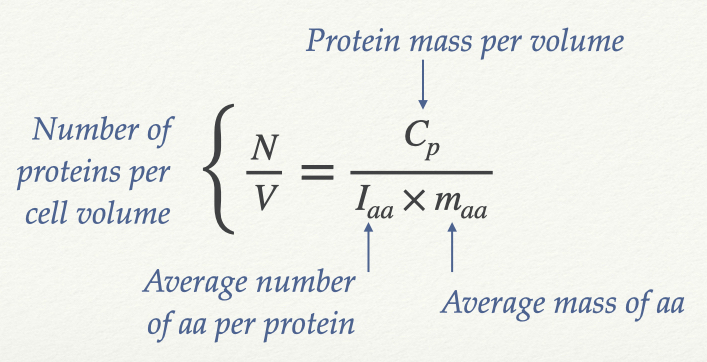
1. You are told that this sequence codes for one gene. Find the longest “gene”, or open reading frame (ORF) corresponding to this DNA sequence. Transcribe this ORF into an RNA sequence and then into its corresponding protein sequence (label the 5’ and 3’ ends of the RNA, and Nter and Cter ends of the protein).
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’-GAGCCATGCA**T**TATCTAGATAGTAGGCTCTGAGAATTTATCT-3’.

What would be the resulting RNA and protein sequences?

**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:



With the following:

* + - * + g/ml
        + Daltons
        + 1 Dalton = 1.66 10-27 kg
        + 1 ml = 1012 m3

Assuming that a bacterium has a volume of 1 m3, how many proteins does this bacterium contain?

**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 |