

# Bio231 - Computational Biology

## Assignment no 1

Assignment Due: February 28<sup>th</sup>, 2022

### Instructions

- Read the section on plagiarism in your student handbook. Plagiarized assignments and copied code will be referred to the DC.
- Assignments must be submitted in the “Assignment” section on LMS as a master zip file containing individual files for Python codes and word file for part 1. Code must be properly commented
- Use **MATLAB** to write code
- Use of built-in libraries and Functions are NOT allowed
- Assignments submitted on Dropbox, emailed assignments or late submissions will not be accepted.
- This is an individual assignment. Discussion is not allowed.

### Part 1:

[10 points]

**Question no 1:** Perform global alignment using the following DNA sequences using dynamic programming.

Sequence 1: **ACTACCAGGGCTATA**

Sequence 2: **ATCACCTAGCTTAA**

Scoring scheme for the alignment is **Match**= +2, **Mismatch**= -3, **Gap**= -4

Note: Perform the necessary traceback step and write all the possible alignments.

### Part 2:

#### Question # 1:

[10 points]

Write a program to load any DNA sequence of your choice from any database. The code should be able to perform following task.

- a. Identify start codons on forward strand
- b. Identify start codons on reverse strand

#### Question # 2:

[10+10 points]

Write a program to translate the DNA sequences having start and stop codon to create amino acid sequences

Note that the DNA sequence translation to amino acid sequence should start from the start codon and end at Stop codon.

#### Question # 3:

[10 points]

Search the amino acid sequences obtained in Question 2 in a protein database to see which reported proteins match those sequences.

**Question # 4:**

[10 points]

Write a program to cut your given sequence at various restriction sites reported in database and identify the patterns of splicing.

**BONUS QUESTION # 5:**

[20 points]

Select 50 random protein sequences.

- a. Find all possible restriction sites
- b. Find patterns of:
- c. Splicing Site
- d. Insulators
- e. Enhancers
- f. Silencers