North East University Bangladesh
Department of Computer Science and Engineering
Semester Final Examination, Summer-2022
Program: BSc.(Engg.) in CSE
Course Code: CSE 449
Course Title: Bioinformatics

Marks: 40   Time: 120 minutes

[NB: FIGURES IN THE MARGIN INDICATE FULL MARKS]
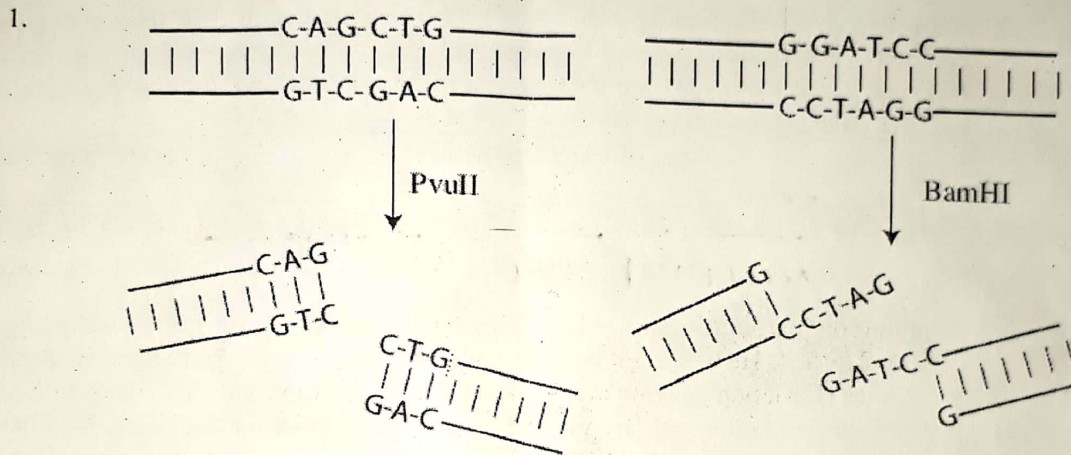
Answer any four questions from the following

1.

Figure 1. Making DNA fragments

Figure 1 describes the sticky ends after cutting DNA with restriction enzymes. BamHI and PvuII cut at $GGATCC$ and $CAGCTG$, respectively, both of which are palindromes.

a) What does restriction mapping mean?   1
b) How do restriction enzymes work?   2
c) What is the first restriction enzyme ever?   1
    i.   BamHI
    ii.   HindI
    iii.   HaeII
    iv.   HindII
d) Define restriction digest and classify it.   1
e) Write the fundamental difference between full and partial digest. Explain with an   3
   example.
f) Explain the partial digest problem in mapping DNA.   2

2. In the partial digest problem, the pairwise distances must have $L = \frac{n(n-1)}{2}$ integers, such that a set of total DNA fragments equals to the L.

a) If you have $n = 5$, then how many DNA fragments are possible for L?   1
b) Consider a set of restriction sites $X = \{2, 4, 0, 10, 7\}$. Find all lengths of DNA   3
   fragments that are possible for X.
c) You are given $\Delta X = L = \{3, 2, 5, 4, 9, 6\}$. Applying the partial digest problem, find the   4
   cutting positions of the DNA map.
d) What do homomeric sets mean for a partial digest problem? Explain it.   2
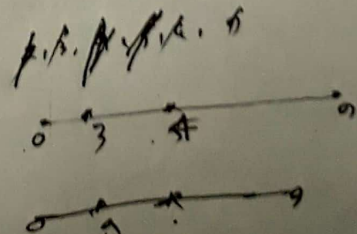
3. Although fruit flies do not have as sophisticated an immune system as humans, they have a small set of *immunity genes* and are always susceptible to infections from bacteria. Most of the time, the genes remain inactive. However, whenever any organs were infected, they switched on and produced proteins to destroy the infection. Biologists set an experiment and found many immunity genes in the fruit fly genome have a pattern or the same strings that are reminiscent of $TCGGGGATTTCC$.

a) In bioinformatics, what is called the pattern of the immunity genes of fruit flies?　　1

b) Describe with an example the representation of the secret string in the immunity genes.　　2

c) If P is a pattern or an available string in all DNA sequences of fruit flies, describe –　　0.5x2=1
　　i.　k-mer
　　ii.　(l, m)

d) Consider a biologist finding a mutation in the t sequences of fruit flies. As you are a computer programmer, the biologist will let you know the starting positions $(s_1, s_2, s_3, ..., s_t)$ of $(7, 2)$ motifs, and you must help her find the implanted motif for the flies. Also, calculate the score.　　6

An array of t DNA sequences	Starting positions, S
AATTCaG TGgGgGA cCGaTTGaGaGAgcT TGgGGgATgCaA TGaGaGAtCgATGaGgGATGGGAcTa	8, 20, 30, 42, 53

4. a) Applying the partial digest problem, construct a homomeric set for the following $L = \{4, 2, 1, 3, 6, 5, 4, 10, 8, 5\}$. If it is impossible, say the X does not produce any multiset.　　5

b) What does restriction enzyme mean? What are the cleavage and sticky ends of DNA?　　1+1=2

c) Without any mutation, you are given t sequences of DNA for fruit flies. You have to find whether any 8-mer secret pattern, P contains in all $t \times n$ matrix or not. Consider n represents the total nucleotides.　　3

4 × 26 matrix
CGGGGATGCAACTTCACATTCCCTTT
TTACGTACGTCAGGTTAAATGCAACT
ATGCAACTACACCATTCCATCTACCT
GTTTGGGACGAATGCAACTGCAGGTA

5. a) Explain the importance of sequence alignment in bioinformatics.　　2

b) What is the purpose of Smith-Waterman and Needleman-Wunsch algorithms in sequence alignment?　　2

c) Find the highest similarity between the nucleotides lab sequence, $LS = AGCT$, and the reference sequence, $RS = ATGCT$. Consider the match, mismatch, and gap penalty for each nucleotide are $+1, -1$, and -2, respectively.　　3

d) Complete the matrix and find the global alignment score between two sequences by applying the appropriate algorithm's initialization, table filling, and traceback process. Consider match $= +1$, mismatch $= -1$, and gap penalty is -2.　　3

		A	T	G	C	T
	0	-2			-8	
A	-2					
G				0		
C			-2			
T						+2