Daffodil International University

Faculty of Science & Information Technology
Department of Computer Science & Engineering
Final Semester Examination, Spring 2025
Course Code: CSE115, Course Title: Introduction to Biology and Chemistry for Computation
Level: 01 Term: 01 Batch: 68

Time: 2:00 Hrs **Marks: 40**

Answer ALL Questions

[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]

1.	A research team is developing a new anti-cancer drug that targets DNA to block the replication of cancer cells. The drug binds directly to DNA's nucleotide bases, interfering with its structure and preventing the cancer cells from multiplying. The team is considering two computational methods: Molecular Mechanics (MM) and Quantum Mechanics (QM) for the drug-DNA interaction. Since the team aims to understand electronic interactions at the binding site, they must choose the most appropriate approach.			CO1
	a)	**Identify** the appropriate computational method for analyzing the binding mechanism of the anti-cancer drug to DNA. **Justify** your answer by comparing both methods, including their governing equations.	[5+5]	
	b)	**Explain** how to calculate the total potential energy in a molecular system. **Describe** the key components involved in the calculation with a relevant example.		
2.	Dr. Emily Roberts, a leading genomics scientist at the Global Antibiotic Resistance Project, is developing an efficient genome alignment algorithm for detecting mutations linked to antibiotic resistance in bacteria. Her team employs the Burrows-Wheeler Transform (BWT) and suffix array-based read mapping to handle the vast amount of sequencing data for efficient DNA sequence alignment. As part of their computational validation, they are given the Burrows-Wheeler Transform (BWT) sequence: "ATTCG$GG". Before proceeding with read mapping, they must reconstruct the original genome and analyze how BWT improves computational efficiency.			
	a)	**Apply** the proper technique to reconstruct the original genome sequence from the given BWT. Provide a detailed step-by-step explanation of the process.	[5+5]	CO2
	b)	**Generate** the Burrows-Wheeler Matrix (BWM) from the original sequence and compute the corresponding Suffix Array (SA(T)). Illustrate the efficiency of BWM and suffix array methods in large-scale genomic sequencing.		
3.	Dr. Rachel is studying the genetic variation in several strains of the Influenza virus to understand how mutations affect its virulence and transmission. He needs to align the entire RNA sequence of the virus strains to understand the differences and similarities between them. He will compare two sequences from different strains and use an alignment strategy that ensures all parts of both sequences are considered. Given Sequences: Sequence 1 (Virus Strain A): AGTACGGA Sequence 2 (Virus Strain B): AAGTAGGA Scoring Scheme: Match: +4, Mismatch: +1 and Gap penalty: -2			CO3
	a)	**Apply** an appropriate alignment algorithm based on the given scoring scheme by constructing the alignment score matrix.	[8+2]	
	b)	**Compute** the optimal alignment and alignment score to find the best alignment between the two sequences over their entire length.		
4.	Mst. Zannatun Nesa, a bioinformatics student, is studying the genetic variations between two species and wants to identify the similarities and differences between their DNA sequences. She is using the			

	FASTA algorithm to compare the sequences by applying hash tables to efficiently match subsequences. She also needs to evaluate the performance of the FASTA algorithm in terms of its sensitivity and selectivity to ensure that the algorithm is accurately detecting significant sequence matches. Query Sequence: ACGGTAGCTA Target Sequence: TCACGGTCT			
	a)	**Apply** the FASTA algorithm with K=1 for both the query sequence and the target sequence to find the best matching subsequences.	**[6+2+2]**	**CO3**
	b)	Calculate sensitivity and selectivity specifically for the base G (Guanine) considering the Query sequence as the dataset and the Target sequence as the outcome.		
	c)	**Interpret** the results based on the sensitivity and selectivity value obtained and assess how well the FASTA algorithm performs in finding the relevant sequence matches between the two sequences.		