Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

End - Term Examinations – December 2025

Date: 15-12-2025 **Time:** 1.00pm to 04.00pm

School: SOCSE	**Program** : CSE	
Course Code: CSE3507	**Course Name:** Algorithms in Computational Biology	
Semester: V	**Max Marks**: 100	**Weightage**: 50%

CO - Levels	**CO1**	**CO2**	**CO3**	**CO4**	**CO5**
Marks	**24**	**24**	**26**	**26**	**----**

Instructions:

(i) Read all questions carefully and answer accordingly.

(ii) Do not write anything on the question paper other than roll number.

Part A

Answer ALL the Questions. Each question carries 2marks. **10Q x 2M=20M**

1.	Outline the historical eras of computational biology and mention a significant breakthrough for each era.	**2 Marks**	**L2**	**CO1**
2.	Classify the core domains of computational biology	**2 Marks**	**L2**	**CO1**
3.	Apply the concept of asymptotic analysis to explain why we compare algorithms based on input size.	**2 Marks**	**L3**	**CO2**
4.	Differentiate between Comparative ImmunoProfiling (CIP) and the Consecutive Integer Problem (CIP) by summarizing four important differences.	**2 Marks**	**L3**	**CO2**
5.	Show how Laplace smoothing can be implemented to avoid zero probabilities when estimating biological data.	**2 Marks**	**L3**	**CO3**
6.	Use exhaustive search to justify its continued importance in complex problem spaces.	**2 Marks**	**L3**	**CO3**
7.	Demonstrate how bounding increases the efficiency of branch and bound in a given scenario	**2 Marks**	**L3**	**CO3**

8.	Apply Bellman's principle of optimality to formulate a dynamic programming solution for a given biological sequence problem.	**2 Marks**	**L3**	**CO4**
9.	Demonstrate four applications of dynamic programming in computational biology	**2 Marks**	**L3**	**CO4**
10.	Illustrate the fundamental concepts involved in the BLAST algorithm, and briefly explain how they contribute to its efficiency in sequence searching.	**2 Marks**	**L3**	**CO4**

Part B

Answer the Questions. **Total Marks 80M**

11.	**a.**	Describe how binary search operates on a sorted list and interpret why its time complexity is more efficient than linear search, using a simple Python example.	**10 Marks**	**L2**	**CO1**
	b.	Summarize how bubble sort is implemented in Python. Interpret how its repeated comparison and swapping mechanism organizes data during sorting.	**10 Marks**	**L2**	**CO1**
Or					
12.	**a.**	Summarize the steps of merge sort implementation in Python and explain how the divide-and-conquer strategy organizes data during sorting.	**10 Marks**	**L2**	**CO1**
	b.	Illustrate the algorithmic approaches applicable to solving major biological problems, and summarize how emerging technologies like machine learning and cloud computing are contributing to this field.	**10 Marks**	**L2**	**CO1**

13.	**a.**	Interpret the concept of asymptotic analysis by describing its main ideas, standard notations, and core properties. Illustrate its use with a relevant example, and explain how it helps evaluate the efficiency of algorithms.	**10 Marks**	**L3**	**CO2**
	b.	Apply the Heap Sort algorithm to a list of gene barcodes by outlining each step in the sorting process. Demonstrate how building a heap and repeatedly extracting the largest element results in a sorted sequence.	**10 Marks**	**L3**	**CO2**
Or					
14.	**a.**	Apply the overlap rule (≥ 5 bases) to the DNA fragments below to join them in Comparative ImmunoProfiling (CIP). F_1: ATGCGTTAGC F_2: TTAGCGGATA F_3: CGGATATCCA	**10 Marks**	**L3**	**CO2**

		F_4: ATTCCAGGGC F_5: AGGGCATGCA Construct the longest possible sequence			
	b.	Interpret the algorithmic steps used in the provided Python program for assembling a genome sequence with greedy TSP-based suffix-prefix overlaps.	**10 Marks**	**L3**	**CO2**

15.	**a.**	Apply the KMP algorithm by writing a program for pattern matching in biological sequences.	**10 Marks**	**L3**	**CO3**
	b.	Apply the Forward and Backward algorithms in Hirschberg's method to solve the Longest Common Subsequence problem for these new sequences: • A=*A*= "TGCATGAC" • B=*B*= "GTACTG" Split string A*A*, show your forward and backward computations, and assemble the final LCS for this case.	**10 Marks**	**L3**	**CO3**
Or					
16.	**a.**	Apply the maximum likelihood method by writing a program to estimate parameters in biological data.	**10 Marks**	**L3**	**CO3**
	b.	Apply graph algorithms by writing a program to construct and traverse a biological interaction graph.	**10 Marks**	**L3**	**CO**

17.	**a.**	Apply the Needleman–Wunsch algorithm to align sequences S = AGT and T = AAGC with match +2, mismatch −1, gap −1. Draw the DP matrix, fill scores, and trace the optimal alignment.	**10 Marks**	**L3**	**CO4**
	b.	Apply the ClustalW algorithm for multiple sequence alignment: describe pairwise distance computation, guide tree construction, and progressive alignment steps.	**10 Marks**	**L3**	**CO4**
Or					
18.	**a.**	Use the Smith–Waterman algorithm and scoring (match +2, mismatch −1, gap −2) to align S = ACACACTA and T = AGCACACA. Build the DP table, fill in scores, and extract the highest scoring local alignment.	**10 Marks**	**L3**	**CO4**
	b.	Apply the UPGMA algorithm for phylogenetic tree construction, explaining its cluster merging logic using distance matrices.	**10 Marks**	**L3**	**CO4**