Name: **Enrolment No:**	

UPES
End Semester Examination, May 2025

Course: Bioinformatics and Computational Biology **Semester: IV**
Program: B.Tech Biotechnology / Biomedical **Time: 03 hrs.**
Course Code: HSBE2006 **Max. Marks: 100**

Instructions: Attempt all questions

S. No.	Section A Short answer questions/ MCQ/T&F (20Qx1.5M= 30 Marks)	Marks	COs
Q 1	The main role of algorithms in computational biology is a) Replacing experimental data b) Designing molecules c) Analyzing and predicting biological behavior d) Coding protein structures	**1.5**	**CO1**
Q 2	The following is NOT a correct application of computational biology a) Vaccine design b) Weather prediction c) Genome annotation d) Protein structure prediction	**1.5**	**CO1**
Q 3	GenBank is maintained by: a) EMBL b) NCBI c) DDBJ d) PDB	**1.5**	**CO1**
Q 4	The Protein Data Bank (PDB) contains information on: a) DNA structures b) RNA folding c) 3D structure of biomolecules d) Sequence motifs	**1.5**	**CO1**
Q 5	Database types that you use to study protein-protein interaction is a) DDBJ b) PDB c) STRING d) BLOSUM	**1.5**	**CO1**
Q 6	Alignment method which is generally used before building a phylogenetic tree is a) Local alignment	**1.5**	**CO2**

	b) Dot plot c) Multiple sequence alignment d) Motif discovery		
Q 7	The algorithm that uses progressive alignment for MSA is a) Needleman-Wunsch b) BLAST c) ClustalW d) FASTA	**1.5**	**CO2**
Q 8	Scoring matrix which is used for aligning distantly related protein sequences is a) PAM30 b) PAM250 c) BLOSUM80 d) BLOSUM100	**1.5**	**CO2**
Q 9	Type of phylogenetic tree which assumes a constant rate of evolution is a) Unrooted tree b) Cladogram c) Molecular clock tree d) Neighbor-joining tree	**1.5**	**CO2**
Q 10	The following is NOT a step in multiple sequence alignment a) Scoring b) Pairwise alignment c) Distance matrix generation d) Ligand fitting	**1.5**	**CO2**
Q 11	Homology modeling is based on the assumption that: a) All proteins are unique b) Sequence similarity implies structure similarity c) Protein folding is random d) No tools are needed	**1.5**	**CO3**
Q 12	Comparative modeling requires: a) Only the amino acid sequence b) A template with known structure c) A drug molecule d) Genomic RNA	**1.5**	**CO3**
Q 13	The purpose of Ramachandran plot is to: a) Predict DNA melting b) Validate protein secondary structure c) Visualize allowable dihedral angles in proteins d) Align RNA molecules	**1.5**	**CO3**
Q 14	AI-based structure prediction tools like AlphaFold improve: a) Docking scores b) Visual rendering c) Sequence homology	**1.5**	**CO3**

	d) 3D structure accuracy of proteins		
Q 15	The prediction of alpha helices and beta sheets from sequence is called: a) Structure-based modeling b) Secondary structure prediction c) RNA folding d) Primary structure alignment	**1.5**	**CO3**
Q 16	Name a pathway database.	**1.5**	**CO4**
Q 17	Expand EMBL	**1.5**	**CO4**
Q 18	Name a database for gene expression data.	**1.5**	**CO4**
Q 19	Provide the principle for homology modeling.	**1.5**	**CO4**
Q 20	What does a phylogenetic tree's "node" represent?	**1.5**	**CO4**
Section B **(4Qx5M=20 Marks)**			
Q 1	List and describe any four applications of bioinformatics in health and agriculture.	**5**	**CO1**
Q 2	Compare and contrast nucleic acid and protein sequence databases with suitable examples.	**5**	**CO2**
Q 3	Discuss how pairwise and multiple sequence alignment helps in understanding evolutionary relationships.	**5**	**CO3**
Q 4	Compare threading and ab initio approaches in 3D protein structure prediction.	**5**	**CO4**
Section C **(2Qx15M=30 Marks)**			
Q 1	Case Study: You are working with a research team investigating a multidrug-resistant strain of *Pseudomonas aeruginosa*. During genome analysis, the team identifies a hypothetical protein (HPX_3214) that shows no significant homology to any known protein in standard databases. However, its expression level increases significantly under antibiotic stress, indicating a potential role in resistance. As a bioinformatics analyst, you are tasked with functionally annotating this hypothetical protein using multiple biological databases. Based on the case study, answer the following questions: 1. Which primary and secondary sequence databases would you consult to gain insights into the hypothetical protein? (4 marks) 2. Describe how domain prediction tools (like InterPro or Pfam) could assist in the functional annotation of HPX_3214. (4 marks)	**15**	**CO3**

	3. Which databases or platforms would help you predict the 3D structure of the protein, and how might that structure inform its function? (4 marks) 4. Explain how a protein-protein interaction (PPI) database could be used to infer the role of this protein in cellular pathways. (3 marks)		
Q 2	Case Study: During an outbreak of a new viral infection, a previously unknown protein from the virus has been sequenced. Structural biologists need your help to predict the 3D structure of this protein to understand its function and interactions with human proteins. Answer the following questions: 1. Explain the process of homology modeling and how it would be used in this case. (3 marks) 2. How would you visualize the predicted 3D structure, and what tools would you use? (3 marks) 3. If no homologs are available, what alternative methods could you employ? (3 marks) 4. Discuss how secondary structure prediction assists in full 3D prediction. (3 marks) 5. What are the limitations of computational structure prediction in this scenario? (3 marks)	**15**	**CO4**
Section D **(2Qx10M=20 Marks)**			
Q 1	A) Differentiate between local and global alignment with examples. B) Describe how dynamic programming is used in the Needleman-Wunsch and Smith-Waterman algorithms.	**4+6**	**CO1**
Q 2	A) Describe how φ (phi) and ψ (psi) angles influence protein folding. Discuss the allowed and disallowed regions, and how the plot helps in validating predicted protein structures. B) Describe the computational approaches for predicting 3D protein structure. How are these protein structures validated and refined.	**6+4**	**CO2**