Name: **Enrolment No:**	

UPES

End Semester Examination, May 2024

Course: Computational Biology and Bioinformatics — **Semester : VI**

Program: B.Sc.+ M.Sc. Microbiology — **Duration : 3 Hours**

Course Code: HSMB2010 — **Max. Marks: 100**

Instructions: Attempt all the questions

S. No.	Section A Short answer questions/ MCQ/T&F (20Qx1.5M= 30 Marks)	Marks	COs
Q 1	Which database contains information about metabolic pathways and associated enzymes? a) GenBank b) UniProt c) KEGG d) GEO	**1.5**	**CO1**
Q 2	Which of the following databases focuses on protein sequences and functional information? a) NCBI b) EMBL c) UniProt d) DDBJ	**1.5**	**CO1**
Q 3	Which database is commonly used for storing information about genetic polymorphisms and mutations? a) GenBank b) dbSNP c) UniProt d) KEGG	**1.5**	**CO1**
Q 4	Expand KEGG	**1.5**	**CO1**
Q 5	What are primary databases. Give an example.	**1.5**	**CO1**
Q 6	Which scoring system assigns scores to aligned residues based on their evolutionary relationship? a) Dynamic programming b) Scoring matrices c) Genetic algorithms d) Hidden Markov models	**1.5**	**CO2**
Q 7	Which matrix series is more suitable for distantly related protein sequences? a) PAM	**1.5**	**CO2**

	b) BLOSUM c) Both are equally suitable d) Neither is suitable		
Q 8	Which of the following is NOT a step in the process of pairwise sequence alignment? a) Scoring the alignment b) Identifying conserved regions c) Extending the alignment d) Calculating evolutionary distances	**1.5**	**CO2**
Q 9	Which type of sequence alignment is particularly useful for identifying conserved functional domains within proteins? a) Global sequence alignment b) Local sequence alignment c) Pairwise sequence alignment d) Multiple sequence alignment	**1.5**	**CO2**
Q 10	Which matrix series is typically used for nucleotide sequence alignment? a) PAM b) BLOSUM c) GATTACA d) None, nucleotide sequences use specific scoring schemes	**1.5**	**CO2**
Q 11	Which of the following organisms has a prokaryotic genome? a) *Saccharomyces cerevisiae* b) *Arabidopsis thaliana* c) *Escherichia coli* d) Human	**1.5**	**CO3**
Q 12	Which of the following is NOT a major feature of the completed genome of Escherichia coli? a) Circular chromosome b) Plasmids c) Mitochondrial genome d) Approximately 4.6 million base pairs	**1.5**	**CO3**
Q 13	Which organism is commonly used as a model organism in molecular and cellular biology? a) *Escherichia coli* b) *Saccharomyces cerevisiae* c) *Arabidopsis thaliana* d) All of the above	**1.5**	**CO3**
Q 14	What is the approximate size of the genome of Saccharomyces cerevisiae? a) 4.6 million base pairs b) 3 billion base pairs c) 12 million base pairs d) 12 billion base pairs	**1.5**	**CO3**

Q 15	Which technique is used to identify and quantify proteins in a biological sample? a) PCR b) Western blotting c) Mass spectrometry d) 2-D gel electrophoresis	**1.5**	**CO3**
Q 16	What is a protein motif? Give an example.	**1.5**	**CO4**
Q 17	What is the term for the 3D arrangement of secondary and tertiary structures in a protein? a) Motif b) Fold c) Domain d) Quaternary structure	**1.5**	**CO4**
Q 18	What technique is used to predict the 3D structure of a protein when experimental data is not available? a) X-ray crystallography b) Nuclear magnetic resonance (NMR) c) Homology modeling d) Cryo-electron microscopy	**1.5**	**CO4**
Q 19	Which region of the Ramachandran plot indicates allowed regions for protein backbone torsion angles? a) Alpha helix region b) Beta sheet region c) Left-handed helix region d) Sterically allowed region	**1.5**	**CO4**
Q 20	What is the primary tool used to determine the tertiary structure of a protein? a) NMR spectroscopy b) X-ray crystallography c) Mass spectrometry d) PCR	**1.5**	**CO4**
Section B (4Qx5M=20 Marks)			
Q 1	Explain the role of file formats like GenBank in standardizing the storage and exchange of nucleic acid sequence data in bioinformatics research.	**5**	**CO1**
Q 2	Explain the importance of scoring matrices in sequence alignment and compare the characteristics of the PAM and BLOSUM series of matrices.	**5**	**CO2**
Q 3	How does 2-D gel electrophoresis contribute to the study of proteomics, and what information does it provide about protein expression patterns?	**5**	**CO3**

Q 4	Explain the process of energy minimization in protein structure refinement and its importance in improving the accuracy and stability of predicted protein structures.	**5**	**CO4**
Section C **(2Qx15M=30 Marks)**			
Q 1	**Case study:** You are a bioinformatics researcher working on a project that involves comparing and analyzing DNA or protein sequences from different organisms or homologous genes. Your goal is to identify conserved regions, detect sequence variations, and infer evolutionary relationships between sequences using pairwise and multiple sequence alignment techniques. Based on your understanding of sequence alignment, answer the following questions A) Discuss how pairwise and multiple sequence alignment can help you in comparing the sequences. B) Explain algorithms for pairwise and multiple sequence alignment. C) List popular bioinformatics tools used for pairwise and multiple sequence alignment. D) Discuss the role of multiple sequence alignment in identification of conserved elements and infer phylogenetic relationship.	**4+4+4+3**	**CO2**
Q 2	**Case study:** You are a microbiologist collaborating with bioinformatics experts on a research project focused on microbial bioremediation of pesticide-contaminated soil. Your team is interested in isolating carbendazim-degrading microbes using a combination of traditional microbiological techniques and advanced bioinformatics tools. Your goal is to identify candidate microbes with the potential to degrade carbendazim and elucidate the genetic mechanisms underlying pesticide degradation. Based on your understanding of microbial degradation, answer the following: A) Explain the methodology and techniques used for isolating and characterizing carbendazim-degrading microbes. B) Discuss the bioinformatics tools and databases available for microbial genomics and metagenomics. C) What metabolic database can be explored for metabolic pathways and enzymatic mechanisms	**4+4+4+3**	**CO4**

	D) Discuss how genome sequencing and data analysis can be helpful in accelerating the discovery		
Section D (2Qx10M=20 Marks)			
Q 1	A) What is 2-D gel electrophoresis? Explain its working principle. B) What factors can affect the resolution and reproducibility of protein spots in 2D gel electrophoresis, and how can they be optimized?	**5+5**	**CO3**
Q 2	A) What is the Ramachandran plot, and how is it used to evaluate the quality of protein structures? B) Describe the interpretation of the Ramachandran plot in terms of allowed and disallowed regions for phi (ϕ) and psi (ψ) torsion angles. C) Discuss the implications of Ramachandran plot analysis for protein structure validation.	**4+3+3**	**CO4**