



M.TECH. DEGREE EXAMINATIONS: APRIL / MAY 2023

(Regulation 2018)

Second Semester

BIOTECHNOLOGY

P18BTI2203: Computational Biology

COURSE OUTCOMES

CO1: Apply knowledge of mathematics and science in biological sequence analysis.

CO2: Analyse and interpret biological sequence data.

CO3: Educate the appropriate selection of tools for protein analysis.

CO4: Analyse and interpret protein interactions.

CO5: Design a bio-based system/ model using artificial neural networks.

CO6: Apply, design and interpret biological data using computational tools.

Time: Three Hours

Maximum Marks: 100

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

- Which of the following will be used to match the protein query against translated nucleotide database? CO2 [K₂]
 - BLASTp
 - tBLASTn
 - tBLASTp
 - BLASTx
- You are given two DNA sequences to align GTCCTTC and GTCTATC CO2 [K₃]

You have a scoring scheme where a

 - match gives you +1
 - a mismatch gives you -1
 - gap costs -1

What is the the best optimal score of the two sequences?

 - 4
 - 6
 - 3
 - 5
- Which branching diagram is assumed to be an estimate of a phylogeny when branching lengths are proportional to the amount of inferred evolutionary change? CO4 [K₃]
 - Phylogram
 - Cladogram
 - A guide tree
 - Cardiogram

4. Match the following

CO3 [K₂]

| List I | List II |
|-----------------------------------|-----------------------|
| A. Secondary Structure Prediction | i. Maximum Likelihood |
| B. Protein Structure Validation | ii. UPGMA |
| C. Tree Evaluation | iii. JPred |
| D. Distance method | iv. PROSESS |

- a) A iv, B ii, C i, D iii b) A iv, B i, C ii, D iii
c) A iii, B iv, C i, D ii d) A ii, B iv, C i, D iii

5. E-value (associated to a score S) is the number of distinct alignments, with a score equivalent to or better than S, that are expected to occur in a database search by chance. The higher the E value, -----

CO1 [K₂]

- a) The more the organisms are related b) The less amount of SNP's and INDEL's were found
c) The less likely the organisms are related d) The more sequence base nucleotides were used

6. Profile Hidden Markov Models (HMMs) are important because they provide a powerful way to search databases for _____ related homologs.

CO4 [K₁]

- a) closely b) distantly
c) New d) Extra

7. Bootstrapping repeats the phylogenetic analysis several times, each time reshuffling the columns of the initial alignment, in order to _____.

CO4 [K₂]

- a) Build a consensus tree, where the number of times each branch reforms is counted and used to estimate its probability. b) Generate a random model from which to benchmark phylogenetic data
c) Produce a graphical representation of the tree. d) Assess the probability that the sequences in the alignment are unrelated.

8. PHD is a method in secondary structure prediction which is based on

CO5 [K₂]

- a) Neural Networks b) Multiple Sequence Alignment
c) Multiple Sequence Alignment and Neural Networks d) Consensus

23. A molecular biologist discovers a new protein sequence. He knows the protein sequence but he is unaware of the structure. In this scenario, explain how he can elucidate the structure of the proteins using computational tools? CO3 [K₅]
24. Examine the different ranges of residue-residue contacts in protein interaction studies CO4 [K₄]
25. Illustrate how Artificial Neural Network plays an important role in Computational Biology CO5 [K₃]
26. Summarize the programs used in molecular modeling. Explain how the model is validated. CO6 [K₄]

Answer any FOUR Questions

PART D (4 x 10 = 40 Marks)

27. Analyze in detail the dynamic programming algorithms used for shotgun DNA sequencing CO1 [K₄]
28. Enumerate the various types of multiple sequence alignment. Give the programs under each category, and mention the drawbacks of progressive alignment. CO2 [K₃]
29. Describe the evolution of protein structure classification, databases and Protein structure prediction with respect to membrane proteins. CO3 [K₃]
30. Discuss how assignment of secondary structures is done in protein structure prediction and explain the process of computation of solvent accessibility. CO4 [K₄]
31. Interpret how Single and multi-layer ANN perception is exploited in gene and secondary structure prediction. CO5 [K₅]
