



M.TECH DEGREE EXAMINATIONS: JUNE 2017

(Regulation 2015)

Second Semester

BIOTECHNOLOGY

P15BTT203: Computational Biology

COURSE OUTCOMES

- CO1:** Understand the concepts of biological data and databases.
- CO2:** Understand sequence alignment methods.
- CO3:** Describe creation and biological motivation for preparing phylogenetic trees.
- CO4:** Understand machine learning techniques as applied to biological data.
- CO5:** Be able to use UNIX and program in Perl.

Time: Three Hours

Maximum Marks: 100

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

1. Assertion (A): β structures are generally rigid than alpha or disordered protein CO4 [K₂]
Reason (R): Low frequency collective vibrations are thought to be sensitive to local rigidity within proteins.
 - a) Both A and R are Individually true and R is the correct explanation of A
 - b) Both A and R are Individually true but R is not the correct explanation of A
 - c) A is true but R is false
 - d) A is false but R is true
2. In a multiple sequence FASTA file format, each new sequence starts with CO2 [K₁]
 - a) *
 - b) -
 - c) #
 - d) >
3. Comparing sequence A to sequence B results an alignment that matches sequences A and B over their whole length, The p-value for the alignment is $<10^{-15}$. Sequence B also has a significant match to sequence C ($p < 10^{-6}$). CO2 [K₄]
 - a) Information is insufficient to identify sequence homology between A and C.
 - b) It's a suggestive. p-value of A and C has to be calculated to compute homology
 - c) Information is sufficient to show A and C are homologous.
 - d) Information is sufficient sequence A and C are heterogeneous.
4. Match List I and List II CO3 [K₂]

List I	List II
A. GENSCAN	i. a phylogenetic analysis tool.
B. PROTPARAM	ii. Predict splicing sites.

C. JIGSAW	iii. Predict exon-intron sites.
D. PAML	iv. Predict physio-chemical properties of protein

A B C D

- a) ii i iii iv
b) iii iv ii i
c) ii iv iii i
d) iii i ii iv

5. Assertion (A): Chou- Fausman method is a empirical technique for the prediction of secondary structures in proteins CO4 [K₃]

Reason (R): The method is based on analyses of the absolute frequencies of each amino acid in alpha helices, beta sheets, and turns based on known protein structures solved with X-ray crystallography

- a) Both A and R are Individually true and R is the correct explanation of A b) Both A and R are Individually true but R is not the correct explanation of A
c) A is true but R is false d) A is false but R is true

6. Two sequences with 95% identity which of the scoring matrices will be the most appropriate CO2 [K₂]

- a) BLOSUM 65 b) BLOSUM 95
c) BLOSUM 50 d) BLOSUM 80

7. Which of the following is not included in bioinformatics? CO1 [K₁]

1. Physical examination of patients
2. Using primary sequence to detect homology
3. Using twitter feed to find trending biological studies
4. Extracting useful information from biological data

- a) 1,3 b) 1,4
c) 1,2 d) 2,3

8. In BLAST search, what does the filter for low complexity do? CO2 [K₂]

- a) It replaces region of low complexity in the databank with the symbol for any residue. b) It replaces regions of low complexity in the query sequence with the symbol for any residue
c) It allows retrieving of "Warning Sequences" that are part of the databank and alerts to the fact that a query is of low complexity d) It trims the regions of low complexity present only in the ends of the query sequence.

9. To perform a non-redundant search with newly sequenced novel protein as query, which search tool would you suggest? CO2 [K₂]

25. Describe a method to evaluate multiple sequence alignment with an example. CO3 [K₂]
26. List the applications and challenges of next generation sequencing over illumina sequencing method. CO1 [K₃]
27. Write code a UNIX script to compare two protein files. CO5 [K₃]
28. Discuss on the objective of aligning sequences with genetic algorithm. CO4 [K₃]
29. Solve using UPGMA Clustering Method and obtain phylogenetic tree for the following data CO4 [K₃]

Species	A	B	C	D
B	7	-	-	-
C	9	11	-	-
D	12	14	6	-
E	15	18	13	4

30. Write a PERL script to transcribe DNA to RNA. CO5 [K₃]

Answer any TWO Questions
PART D (2 x 10 = 20 Marks)

31. Consider that a scientist working on a media optimization problem. Four key parameters have been identified pH, Temperature, Concentration and water content. Considering the above parameters, construct a neural network model. Give a note on the mathematical function utilized in the hidden layer. CO4 [K₅]
32. A) Justify the statement “Comparative genomics as a tool to understand evolution and disease”. Support your answer with any case study elements (5) CO4 [K₄]
B) A novel protein is been identified and sequenced. Performing a BLAST search with the FASTA sequence revealed four homologous proteins with almost 78% identify. Explain how 3D structure is predicted from the homologous protein structures. Add a note on the structural verification of 3D protein structures. (5) CO2 [K₄]
33. Align the following pentide sequences using smith-waterman algorithm under the following base conditions: CO2 [K₄]
Sequence 1 : A W R G P H Q R and
Sequence 2 : S W G P Q
M=+5; Ms=-2 and Gap = -5.
Comment about the similarity of the sequences.
