Q. Code: 461627

M.E. / M.TECH. DEGREE EXAMINATIONS, DEC 2020 (Held during April, 2021)

First Semester

BY18102 – Computational Systems Biology

(Regulation 2018)

Time: Three hours Maximum : 80 Marks

Answer **ALL** questions

PART A - (8 X 2 = 16 marks)

1. What is the difference between Refseq and Genbank?

   a) RefSeq includes publicly available DNA sequences
   b) GenBank includes nonredundant curated data
   c) GenBank sequences are derived from RefSeq
   d) RefSeq sequences are derived from GenBank

2. Which of the following is not true about secondary protein structure?

   a) The hydrophilic/hydrophobic character of amino acid residues is important to secondary structure

   b) The ability of peptide bonds to form intramolecular hydrogen bonds is important to secondary structure

   c) The alpha helix, beta pleated sheet and beta turns are examples of protein secondary structure

   d) The steric influence of amino acid residues is important to secondary structure

3. Which of the following is incorrect about a microarray?

   a) It is a slide attached with a high-density array of immobilized DNA oligomers representing the entire genome of the species under study

   b) Array of immobilized DNA oligomers cannot be cDNAs

   c) Each oligomer is spotted on the slide and serves as a probe for binding to a unique complementary cDNA

   d) It is the most commonly used global gene expression profiling method

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

   a) Closely b) Distantly c) New d) Extra

5. Name any two secondary structure prediction tools.

6. List the various rendering tools in structure visualization.

7. Give any two applications of decision tree in computational biology.

8. Mention about Hash data type in PERL with an example.

PART B - (4 X16 = 64 marks)

9. (a) Construct the dynamic programming Matrix and hence find the global alignment between the following nucleotide sequences: **(16)**

   Seq1: CTCGCAGC

   Seq2: CATTCAC

   Scoring function: match 10, mismatch -2, gap penalty -5.

(OR)

(b) (i) Define E-value in BLAST program. Describe the various steps involved in BLAST. List the various BLAST algorithms utilized to analyze nucleotide and protein sequences. **(8)**

(ii) Write a short note on Next Generation Sequencing Technique. Give an emphasis on nanopore sequencing and Illumina sequencing method. **(8)**

10. (a) Solve the UPGMA clustering method and obtain phylogenetic tree for the following data: **(16)**

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

(OR)

(b) Write a program in PERL to accept a protein sequence and find its length and the count of each of the aromatic aminoacids Y, F and W in the sequence. **(16)**

11. (a) (i) Propose the steps in homology modeling when the sequence identity between the target and template is <30%. **(8)**

(ii) How would you assess whether a lead compound is an effective antagonist of a certain receptor? Discuss the methodology with the appropriate tools. **(8)**

(OR)

(b) Explain the principle and steps involved in Molecular dynamic simulations with suitable examples. **(16)**

12. (a) Define Machine learning approach? Describe the steps involved in training an Artificial Neural Network. Mention its application in protein structure prediction. **(16)**

(OR)

(b) (i) Describe the role of EST's in genome annotation. **(8)**

(ii) Mention the applications of informatics techniques in DNA computing. **(8)**