B.TECH/BT/5TH SEM/BIOT 3102/2016

BIOINFORMATICS
(BIOT 3102)

Time Allotted : 3 hrs **Full Marks : 70**

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and any 5 (five) from Group B to E, taking at least one from each group.

Candidates are required to give answer in their own words as far as practicable.

Group – A
(Multiple Choice Type Questions)

1. Choose the correct alternative for the following: **10 × 1 = 10**

(i) In a relational database, data is organized in a set of tables. Each table is also known by which of the following terminologies?
(a) A relation (b) A query
(c) A value (d) An attribute.

(ii) The core data in the UniProt database does NOT contain
(a) sequence clusters (b) sequence archive
(c) protein knowledgebase (d) protein structures.

(iii) An ORF does NOT contain
(a) Coding region (b) Start codon
(c) Stop codon (d) Transmembrane helices.

(iv) Cross-validation in QSAR is indicated by which of the choices below?
(a) r^2 is negative (b) r^2 is < 0.5
(c) r^2 is = 0.5 (d) r^2 is > 0.5.

(v) An example of an energy minimization algorithm is
(a) steepest descent (b) OPLS
(c) dielectric constant product (d) MOPAC.

(vi) Which of the following is true about PAM matrices?
(a) Created by Chow-Fasman.
(b) Created using conserved DNA families
(c) PAM-1 means 1 accepted point mutation per 100 residues
(d) Lower numbered PAM matrices are appropriate for comparing distantly related species

(vii) The meaning of E-value in BLAST is
(a) the probability that the query sequence and the subject sequence come from the same organism.
(b) the probability that the query sequence and the subject sequence are homologous.
(c) the expected number of generated sequences that would have the observed alignment (or better).
(d) the inverse of the similarity between the query sequence and the subject sequence.

(viii) Which of the following is a valid scalar variable in PERL?
(a) dna1seq (b) dna!seq
(c) 1dnaseq (d) !dnaseq.

(ix) To find a capital letter after any but possibly no spaces at the start of the string
(a) \s[A-Z] (b) \s.*[A-Z] (c) \s*[A-Z] (d) ^.\s\U.

(x) BLAST X program performs BLAST
(a) translating a protein sequence
(b) translating the DNA database
(c) translating the input sequence
(d) none of these.

Group – B

2. (a) Using classical definitions, state the differences between bioinformatics and computational biology? Write the application of biocomputing tools in areas of sequence analysis.

(b) What are the two limitations of bioinformatics analyses output? Why is systems level simulation and integration considered the future of bioinformatics?

(4 + 2) + (3 + 3) = 12

3. (a) What are the type of database structures that are used in biological databases? Differentiate between primary, secondary and tertiary biological databases. Cite and briefly explain three problems that are associated with biological databases and steps taken to solve them.

(b) What are the reasons for development of a protein structural classification system? Itemwise describe the basic characteristics of CATH database with examples. Explain, with examples, three major characteristics of NCBI Gene sub-database.

(2 + 2 + 2) + (2 + 2 + 2) = 12

Group – C

4. (a) In the context of similarity, describe the following items with suitable graphical representation: safe zone, twilight zone and midnight zone.

   (b) Briefly describe the steps of the BLAST algorithm.

   (c) Cite the use of adjustable gap penalties in Clustal. What is the name of the statistical indicator in BLAST result*?* How is it related to the raw alignment score? What is the formula?

3 + 4 + (2 + 1 + 1 + 1) = 12

5. (a) What are the differences between prokaryotic and eukaryotic gene finding approaches**?**

   (b) Define the four basic parameters used for nucleotide prediction accuracy. What are the mathematical formulae for Sn and Sp? Interpret the formulae. What is the mathematical formula for the correlation coefficient (CC) used in gene finding? What is its range and meaning?

4 + (2 + 2 + 2 + 2) = 12

Group – D

6. (a) What are the different types of variables used in PERL?

   (b) Mention the use of "tr" operator in PERL program.

   (c) Write a program to report how "GC rich" a particular sequence is. [Hint: Percentage of G and C in the DNA].

   (d) Using the same sequence, write another PERL program to find out its complementary sequence.

3 + 1 + 4 + 4 = 12

7. (a) Write a subroutine to concatenate two strings of DNA.

   (b) Write a program in PERL to determine the frequency of nucleotides; assume that the data will be taken from a file.

6 + 6 = 12

Group – E

8. (a) Use an example to set up a QSAR equation for an inhibitor drug binding to a drug target. Define the relevant parameters in the equation and what are the implications of such a correlation?

   (b) Draw the hypothetical potential energy landscape of a protein with proper labels

   (c) Write down the general conformational energy expression for a protein explaining all the terms in the equation.

4 + 3 + 5 = 12

9. (a) Draw a flowchart to represent the steps in a homology modelling algorithm for the tertiary structure of a globular protein.

   (b) Why are the template identification and loop modelling steps the most critical and computationally intensive in 3D structure prediction algorithms?

   (c) Use a diagram to depict the steps in the pairwise energy approach for protein folds prediction.

4 + 4 + 4 = 12