B.TECH/BT/5TH SEM/BIOT 3102/2024

BIOINFORMATICS
(BIOT 3102)

Time Allotted : 2½ hrs **Full Marks : 60**

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and any 4 (four) from Group B to E, taking one from each group.

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

Group – A

1. Answer any twelve: **12 × 1 = 12**

Choose the correct alternative for the following

(i) Which output from a BLAST search provides an estimate of the number of false positives from a BLAST search?
(a) E value (b) Bit score
(c) Percent identity (d) Percent positives

(ii) Which of the following does not describe BLOSUM matrices?
(a) It stands for BLOcks Substitution Matrix
(b)It was developed by Henikoff and Henikoff
(c) The year it was developed was 1992
(d) These matrices are logarithmic identity values.

(iii) Gaps are added to the alignment because it _______
(a) increases the matching of identical amino acids at subsequent portions in the alignment
(b) increases the matching of dissimilar amino acids at subsequent portions in the alignment
(c) reduces the overall score
(d) enhances the area of the sequences.

(iv) To test whether perl is in your PATH, use_
(a) perl-I (b) perl-l (c) perl-el (d) perl-e

(v) Which of the following expression matches a line which begins with a?
(a) /^a/ (b) /@a/ (c) /$a/ (d) /~a/.

(vi) Which of the following escape sequence represents carriage return?
(a) \r (b) \n (c) \n\r (d) \c.

(vii) Which among the following residues is most likely to be present on the surface of a protein?
(a) Val (b) Leu (c) Ile (d) Arg.

(viii) The peptide group in a protein has a rigid, planar structure. Which interactions give the peptide bond its partial double bond character?
(a) Covalent (b) Non-covalent (c) Hydrophobic (d) Resonance.

(ix) What are the most common regular secondary structures found in proteins?
(a) Alpha-helix and turns (b) Beta-sheets and loops
(c) Loops and turns (d) Alpha-helix and beta-sheets.

(x) Which of the following approach is considered under the 'Ligand based drug designing'?
(a) Molecular docking (b) Pharmacophore modeling
(c) QSAR Modeling (d) (b) and (c) both.

Fill in the blanks with the correct word

(xi) If two sequences are descended from a common evolutionary origin, they are said to____________.

(xii) __________ is a graphical way of comparing two sequences in-a two dimensional matrix.

(xiii) For both protein and DNA sequences, there may be regions that contain highly repetitive residues in sequence called as___.

(xiv) The computational methodology that tries to find the best matching between two molecules, a receptor and ligand is called __________ .

(xv) The structure formed by joining the amino acids by a peptide bond is called ________ structure of a protein.

Group - B

2. What are the two databases that typically serve as repositories for structural classification of proteins? Illustrate the categories of data in these two databases.
[(CO1)(Analyze/IOCQ)]
(4 + 8) = 12

3. (a) Using examples of biological databases and portals, explain the differences between flat file, RDBMS and OODBMS formats. Use one of your examples to explain the need for a transition from one format to another in database construction. *[(CO1)(Remember-Understand/LOCQ)]*
(b) How would you relate between structure and function analysis in Biocomputing tool development? Give intrinsic connections between the two as examples.
[(CO1)(Remember-Understand/LOCQ)]
(c) What role does bioinformatics play in computational studies of protein-ligand interactions? *[(CO1)(Remember-Understand/LOCQ)]*
(4 + 3) + 3 + 2 = 12

Group - C

4. (a) Out of the two protein sequences of catalase one sequence is of *Vibrio cholarae* and another of *Drosophilla melanogastor*, justify which kind of pairwise algorithm you will choose to find out the sequence similarity among them. *[(CO2) (Consider/IOCQ)]*

   (b) In performing alignment of protein sequences, the parameters considered are match/mismatches as also the similarities between amino acids. How are these parameters represented in practice? Protein alignments also make use of various substitution matrices. Justify the conditions under which one type of substitution matrix is preferred over the other. *[(CO2)(Evaluate/HOCQ)]*

   (c) What is the basic problem solving mechanism that is incorporated in the BLAST algorithm? Explain the steps of Iterated BLAST. Itemize the advantage(s) of an iterated BLAST search compared to a simple BLAST search. *[(CO2)(Explain/LOCQ)]*

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

5. (a) Itemize the distinguishing features of ab initio based approaches of gene prediction. *[(CO3)(Justify/IOCQ)]*

   (b) (i) For prokaryotic gene prediction, instead of identification of the initiation codon, other features associated with translation initiation may be used-explain how such an approach is helpful for accurate gene prediction. *[(CO3)(Justify/HOCQ)]*
   (ii) Pointwise explain the use of the TESTCODE method for this purpose. *[(CO3)(Justify/IOCQ)]*

   (c) Briefly and graphically represent the three states in a hidden Markov model (HMM) using a labelled diagram. *[(CO3)(Remember/IOCQ)]*

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

Group - D

6. (a) Write a small program using Python to remove the first n characters from a string. *[(CO4)(Describe/LOCQ)]*

   (b) Write a small program using Python to print characters from a string that are present at an even index number. *[(CO4)(Describe/LOCQ)]*

   (c) Explain the syntax of Python functions with a small example. *[(CO4)(Describe/LOCQ)]*

4 + 3 + 5 = 12

7. (a) What are regular expressions in Perl? *[(CO4)(Describe/LOCQ)]*

   (b) How are regular expressions used and what are some common applications of regular expressions? *[(CO4)(Describe/LOCQ)]*

   (c) Write a program in Perl showing the use of the following: convert a string to an array and vice versa. *[(CO4)(Describe/LOCQ)]*

3 + (3 + 2) + 4 = 12

Group - E

8. (a) The secondary structure of an unidentified protein of sequence length 525 amino acids has been predicted by a second generation information theory based algorithm to be 35% α-helix, 20% β-sheet and 45% random coil. Another

comparable prediction protocol using a 3rd generation artificial neural network algorithm yielded 45% α-helix, 23% β-sheet and 32% random coil. Use and interpret the above results to answer the following questions:

(i) Why is a neural network based prediction considered to be more accurate than 2nd generation algorithms?

(ii) What are the physico-chemical factors that are responsible for the formation of α-helices and β-sheets?

(iii) Based on your knowledge, how does your answer impact your choice of algorithm for doing the secondary structure prediction of the particular protein? *[(CO5)(Analyze/IOCQ)]*

(b) Cite the three methodologies of doing tertiary structure prediction of proteins briefly explaining the three methods. Compare the steps in two of these methods that are procedurally similar. *[(CO6)(Remember-Understand/LOCQ)]*

(2 + 2 + 3) + 5 = 12

9. (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? *[(CO6)(Understand-Analyze/LOCQ]*

(b) Draw the hypothetical potential energy landscape of a protein with proper labels. *[(CO6)(Understand-Analyze/LOCQ]*

(c) Write down the general conformational energy expression for a protein explaining all the terms in the equation. *[(CO6)(Remember/LOCQ)]*

4 + 3 + 5 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	60.41	31.25	8.34