

**BIOINFORMATICS
(BTC3102)**

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 of the following is a structural database?
(a) CATH (b) SCOP
(c) VAST (d) PDB
- (ii) Which of the following is INCORRECT about ENTREZ?
(a) It provides a series of forms that can be filled out to retrieve a Medline reference related to the molecular biology sequence databases.
(b) It is a resource prepared only by the staff of the National Center for Biotechnological Information.
(c) ENTREZ provides a straightforward way to access the sequence databases
(d) Effective use of ENTREZ requires and understanding of integrated search features
- (iii) For Needleman-Wunsch algorithm which of the following statement is not an advantage?
(a) New algorithmic improvements as well as increasing computer capacity.
(b) Make it possible to align a query sequence against a large DB in a few minutes.
(c) This does not help in determining evolutionary relationship.
(d) If you have two genes that are well understood as closely related, then this type of algorithm can be used to understand them in further details.
- (iv) Which of the following sentences best describes the difference between a global alignment and a local alignment between two sequences?
(a) Global alignment is usually used for DNA sequences, while local alignment is usually used for protein sequences.
(b) Global alignment has gaps, while local alignment does not have gaps.
(c) Global alignment finds the global maximum, while local alignment finds the local maximum.
(d) Global alignment aligns the whole sequence, while local alignment finds the best subsequence that aligns.

- (v) 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
- (vi) In regular expressions, which of the following pair of pattern is wrongly matched with its significance?
 (a) [] -Or (b) { } -Not
 (c) () - Repeats (d) Z -Any
- (vii) State the correct one of the following expression matches a line which begins with a?
 (a) /^a/ (b) /@a/
 (c) /\$a/ (d) /~a/
- (viii) Which among the following structure of a protein describes the folding of its secondary structural elements and specifies the position of every atom in a protein, including those of side chains?
 (a) Primary structure (b) Secondary structure
 (c) Quaternary structure (d) Tertiary structure
- (ix) 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
- (x) Which one method used for virtual screening
 (a) ADMET analyses (b) QSAR modeling
 (c) Pharmacophore modelling (d) All of the above

Fill in the blanks with the correct word

- (xi) If two sequences are descended from a common evolutionary origin, they are said to____.
- (xii) For both protein and DNA sequences, there may be regions that contain highly repetitive residues in sequence called as _____.
- (xiii) Python was developed by_____.
- (xiv) The identification of drugs through the gene study is called _____.
- (xv) _____ chemical database is maintained by European Bioinformatics Institute.

Group - B

2. (a) What are the characteristics of biological databases? What is the nature of data access for these two databases?
 (b) What are the full forms of SCOP and CATH? Mention characteristics of them.

[[CO1](Analyze/IOCQ)]

[[CO1](Remember-Understand/LOCQ)]

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

3. (a) How is quality control undertaken in a biological database? Using PDB as an example, explain two in-built data quality tools that are present in this protein structure database. *[[CO1)(Analyze/IOCQ]]*
- (b) List the possible inconsistencies that are present in a biological database. Cite three relevant methods by which reorganization of data in a database is typically done. *[[CO1)(Analyze/IOCQ]]*
- (3 + 3) + (3 + 3) = 12**

Group - C

4. (a) There are two sequences one of bacteria another of a plant origin of a particular protein source. In order to find relationship between them which algorithmic approach would you choose? Explain your answer. State which substitution matrix will play a pivotal role here. *[[CO2)(Argue/HOCQ]]*
- (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]]*
- (1 + 2 + 1) + (2 + 2) + (2 + 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 markovmodel (HMM) using a labelled diagram. *[[CO3)(Remember/IOCQ]]*
- 4 + (2 + 2) + (2 + 2) = 12**

Group - D

6. (a) Mention three major characteristics of PERL and PYTHON. *[[CO4)(Describe/LOCQ]]*
- (b) Write PERL programs where you are asked to read the protein sequence from a specific file: protseq.pep. *[[CO4)(Design/IOCQ]]*
- 6 + 6 = 12**
7. (a) Explain the concept of arrays in Perl. *[[CO4)(Describe/LOCQ]]*
- (b) How are arrays used? What are some common array operations? *[[CO4)(Describe/LOCQ]]*
- (c) How are variables declared in Perl? *[[CO4)(Describe/LOCQ]]*

- (d) Write a program showing the following actions: add or remove elements from an array.

[[CO4](Design/IOCQ)]

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

Group - E

8. (a) A computational approach towards studies of protein-ligand interactions can be done through quantitative analysis of relevant binary physico-chemical interactions and bioinformatics based procedures." Explain this statement using the simple example of a protein P binding to a ligand L forming a binary adduct P-L. Your answer should be based on quantitative logic and analysis.
[[CO6](Understand-Apply/IOCQ)]
- (b) Write out the steps of a generalized homology modelling based method that is used for the tertiary structure prediction of a protein. [[CO5](Understand-Analyze/IOCQ)]
- (c) Choose a specific bioinformatics tool that is used for this purpose: What was/were the primary algorithm(s) used? [[CO5](Understand-Analyze/IOCQ)]
- (d) How has this algorithm been used for the purpose of annotating a gene?
[[CO5](Understand-Analyse/IOCQ)]
- $$6 + 3 + 1 + 2 = 12$$
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	37.5	52.08	10.42