ADVANCED BIOINFORMATICS
(BIOT 5201)

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

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and any4 (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) A gap opening penalty for any gap (g) and a gap extension penalty for each element in the gap (r) are most often used, to give a total gap score wx, according to the equation _____
(a) wx – rx = -g (b) wx = g – rx
(c) wx = g + rx (d) wx + g + rx = 0

(ii) 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.

(iii) Which of the following doesn't describe PAM matrices?
(a) This family of matrices lists the likelihood of change from one amino acid to another in homologous protein sequences during evolution
(b) There is presently no other type of scoring matrix that is based on such sound evolutionary principles as are these matrices
(c) Even though they were originally based on a relatively small data set, the PAM matrices remain a useful tool for sequence alignment
(d) It stands for Percent Altered Mutation.

(iv) Which of the following is not a variant of BLAST?
(a) BLASTN (b) BLASTP (c) BLASTX (d) TBLASTNX

(v) 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.

(vi) Which among the following structure of a monomeric 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.

(vii) Lipinski's rule of five is used for
(a) Docking (b) Similarity search
(c) Drug likeness (d) Dynamics simulation

(viii) What are the pitch length of alpha-helix and the number of residues per turn?
(a) 5.3 Armstrong, 6.4 residues (b) 3.4 Armstrong, 5.6 residues
(c) 3.6 Armstrong, 5.4 residues (d) 5.4 Armstrong, 3.6 residues.

(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 of these is gene prediction algorithm?
(a) UPGMA (b) Hidden Markov Model
(c) Maximum parsimony (d) None.

Fill in the blanks with the correct word

(xi) The statistical analysis of alignment scores is much better understood for _____ alignment than for _____ alignment.

(xii) In a phylogenetic tree, the connecting point where two adjacent branches join is called a _____.

(xiii) The peptide group in a protein has a rigid, planar structure. _____ interactions give the peptide bond its partial double bond character.

(xiv) _____ is a structural database.

(xv) _____ alignments are more commonly used for dissimilar sequences with suspected regions of similarity.

Group - B

2. (a) Mention what do you mean by content sensor and signal sensor in respect to gene prediction programme. *[(CO3)(Analyse/HOCQ)]*
(b) Ab-initio based approaches in eukaryotic system of gene prediction programs rely on the several features - Describe them. *[(CO2)(Understand/LOCQ)]*
(c) Mention the characteristics of eukaryotic gene content sensor citing one example of such software used for this purpose. *[(CO2)(Apply/IOCQ)]*

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

3. (a) Write the basis of ab initio based gene prediction. *[(CO4)(Understand-LOCQ]*
(b) Summarise the role of different gene signals associated with this approach. *[(CO4)(Evaluate-HOCQ)]*
(c) 'Presence of mere start codon is sufficient to initiate the beginning of the frame of translation initiation'- State whether the above statement is true or false in respect to bacterial gene prediction. Justify your answer with suitable example. *[(CO4)(Evaluate-HOCQ)]*

(d) 'Prediction program does not predict accurately; performance evaluation is needed for the assessment study'-Evaluate the performance evaluation process.

[(CO4)(Evaluate-HOCQ)]

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

Group - C

4. A, B, C, D are four taxa whose distances are given: AB=0.40, AC=0.35, AD=0.60, BC=0.45, BD=0.70 and CD=0.55 based on suitable alignment method construct the phylogenetic tree. Show step wise how the final tree is developed. Name one bioinformatics software tool that is based on the clustering method you adopted. Enumerate the advantages and disadvantages of the method. *[(CO3)(Construct-IOCQ)]*

(6 + 1 + 2 + 3) = 12

5. (a) 'The effect of homoplasy can be corrected by virtue of using some standard models'- define homoplasy. *[(CO3)(Evaluate/HOCQ)]*

(b) Suppose during this calculation it is assumed all the mutation which happened in sets of the sequences are substituted with equal probability-justify if it is a correct or incorrect assumption and based on that mention briefly the model that will be chosen for further processing. *[(CO3)(Evaluate/HOCQ)]*

(c) To overcome rate heterogeneity among sites what approach will be followed to improve the estimation of evolutionary distances? Justify this with the help of graphical representation. *[(CO3)(Examine/HOCQ)]*

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

Group - D

6. (a) Use an algorithm to explain how a secondary structure of a protein can be predicted using multiple sequence alignment and neural networks.

[(CO4,5)(Apply/IOCQ)]

(b) Tabulate and differentiate between the two major protein structure classification system. *[(CO4,5)(Apply/IOCQ)]*

6 + 6 = 12

7. (a) If an *ab-initio* algorithm was used to predict the tertiary structure of a protein, write out the steps of this algorithm. *[(CO4,5)(Apply/IOCQ)]*

(b) Write the differences in the methodologies of fold recognition and homology modeling. *[(CO4,5)(Apply/IOCQ)]*

6 + 6 = 12

Group - E

8. Describe briefly the conditions for structure based and ligand based drug design Illustrate with example for each. *[(CO5)(Analyse/HOCQ)]*

(6 + 6) = 12

9. (a) Explain the step of energy minimization with respect to protein ligand binding.. *[(CO5)(Analyse/IOCQ)]*

(b) Describe Metropolis Monte Carlo algorithm with respect to a protein optimization problem. *[(CO5,6)(Describe/LOCQ)]*

(c) Explain simulated annealing. *[(CO6)(Remember/IOCQ)]*

4 + 6 + 2 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	12.5	47.9	39.5

Course Outcome (CO):

After the completion of the course students will be able to

1. Use acquired knowledge about different bioinformatics experiment categories (e.g. sequence, structure analysis) and their applications in new biology) (e.g. genomics, proteomics)
2. Learn organization and characteristics of primary and specialized databases and portals, introduction to new applications of databases/portals towards study of metabolic pathways and systems biology
3. Learn and apply sequence alignment methodologies (including comparison of applicable heuristic and dynamic algorithms) for pairwise and multiple sequence alignment and molecular phylogenetics
4. Learn and apply bioinformatics based software tools (and the algorithms underlying them) for annotation and structure prediction of prokaryotic and eukaryotic genes, RNA secondary structure prediction and secondary structure prediction of globular, fibrous and membrane proteins (e.g. use of artificial neural network and Hidden Markov model based algorithms for these purposes)
5. Principles and applications of homology, fold recognition, and ab initio based algorithms for tertiary structure prediction of proteins, application of protein tertiary structure prediction towards problems of protein folding and design.
6. Learn and apply the principles of molecular modelling and energy minimization for small molecule -protein and protein-protein binding; learn the principles and methodologies of computer aided drug design.

**LOCQ: Lower Order Cognitive Question; IOCQ: Intermediate Order Cognitive Question; HOCQ: Higher Order Cognitive Question.*