2016

# ADVANCED BIOINFORMATICS
## (BIOT 5201)

**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 alternatives for the following: **10 × 1=10**

(i) Target selection in drug discovery necessitates
- (a) linking a target molecule that affects disease by affecting its function or expression
- (b) finding information about specific protein-protein interactions
- (c) knowledge about differential genomics and proteomics
- (d) all of the above.

(ii) A Which is a data retrieving tool?
(a) ENTREZ (b) EMBL (c) PHD (d) all of these.

(iii) A kissing hairpin refers to a
- (a) hydrogen bond
- (b) hairpin-bulge contact
- (c) hydrogen bonded interaction formed between loop residues of two hairpin structures
- (d) all of the above

(iv) CLUSTALW does
(a) Local alignment (b) Global alignment
(c) Partial alignment (d) Multiple sequence alignment

(v) A Markov model describes
(a) a sequence of events that that occur consecutively in a chain
(b) a sequence of events that occur in an arithmetic progression in a chain
(c) a sequence of events that occur in a geometric progression in a chain
(d) none of the above.

(vi) An open reading frame (ORF) begins with
(a) an ATG initiation codon (b) a stop codon
(c) an expressed sequence tag (d) a TATA box.

(vii) Which of the following is NOT a secondary structure prediction algorithm?
(a) ZPRED (b) PHD
(c) DSSP (d) SWISS-MODEL.

(viii) Principal component analysis (PCA) is usually used to
(a) filter redundant descriptors
(b) filter essential descriptors
(c) transfer uncorrelated vehicles
(d) all of the above

(ix) Which of the following is true about PAM matrices?
(a) Created by Chou-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

(x) 10 Alignment of two sequences is performed using
(a) dot matrix analysis
(b) dynamic programming algorithm
(c) wordor K-tuple method
(d) all of these

## Group - B

2. (a) In the context of similarity, mention the relationship of the following items citing definitions with suitable graphical representation: safe zone, twilight zone and midnight zone.

(b) Briefly describe steps of BLAST algorithm. Cite the use of adjustable gap penalties in CLUSTAL.

(c) Mention the name of the statistical indicator in BLAST result and mention how it is related to raw alignment score. Also mention the formula.

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

3. (a) Briefly describe the process of computational gene finding.

(b) What are the approaches of computational gene annotations? Cite one example from each category.

(c) Why is *ab initio* method said to be not suitable for eukaryotic gene prediction?

(d) To evaluate the accuracy of the predicted programs mention some parameters which are important. Also state how they help in the evaluation process.

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

## Group - C

4. (a) Cite one difference between phylogram and cladogram.

(b) Find out the number of rooted trees that can be generated from six taxa.

(c) Mention the role of biological data for constructing molecular phylogenetic tree.

(d) Elucidate the role of synonymous and nonsynonymous substitutions in understanding the evolutionary process.

(e) Write the objective of using Kimura substitution model over Jukes-Cantor Model

(f) Calculate the corrected evolutionary distance by suitable substitution model of two sequence where they differ by 30%, out of which 20% of changes are a result of transitions and 10% of changes are a result of transversions.

**2+2+2+2 +2+2 = 12**

5. (a) Using the suitable phylogenetic tree describe the following terminologies:
clade , lineage and paraphyletic group.

(b) Draw an assumptive tree from the following Newick format for tree representation providing suitable reasons:
(((X:1,Y:2),Q:2),(Z:1.5,L:3))

(c) Describe the term 'among site rate heterogeneity' and elucidate how these artifacts can be corrected in the evolutionary models.

(d) 'Bootstrapping is a statistical method used for evaluation of reliability of the inferred phylogeny' – justify the statement with suitable reasons.

**4 + 3 + 3 + 2 = 12**

## Group - D

6. (a) How is helical membrane protein prediction undertaken?

(b) What are the special characteristics of Phobius?

(c) How is tertiary structure prediction undertaken in transcription regulatory proteins containing leucine zippers?

**4 + 4 + 4 = 12**

7. (a) Define a neural network.

(b) Explain its method of operation with respect to protein secondary structure prediction.

(c) Draw a schematic diagram of a protein secondary structure prediction algorithm that uses both multiple alignments and neural networks.

(d) What improves the accuracy of a secondary structure prediction algorithm using a combination of factors? Explain your answer.

**2 + 3 + 4 + 3 = 12**

## Group - E

8. (a) Define QSAR .

(b) What are some typical applications of QSAR?

(c) What are the assumptions of QSAR techniques?

(d) Name two types of descriptors that are used in QSAR studies.

**2 + 3 + 4 + 3 = 12**

9. (a) What are the two main approaches in computer assisted drug design? Define a binding site in a protein. Specify a wet experimental method of identifying a binding site in a protein.

(b) Define molecular docking including its essential characteristics. What is the purpose of a scoring function? Name the factors that affect the docking score explaining two of them.

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