

(Following Paper ID and Roll No. to be filled in your Answer Book)

PAPER ID : 9592

Roll No.

B.Tech.
(SEMESTER. IV) THEORY EXAMINATION, 2012-13
BIOINFORMATICS-I

Time : 3 Hours]

[Total Marks : 100

SECTION – A

1. Attempt all question parts : **10 × 2 = 20**
- (a) Name few path-way databases.
 - (b) Mention two other types of substitution matrix than PAM and BLOSUM.
 - (c) Give the advantage and disadvantage of shotgun sequencing method.
 - (d) What is sequence filtering ? How can you identify compositional complexity of a sequence ?
 - (e) How to calculate the significance of a local alignment using Gumbels extreme value distribution ?
 - (f) What are the three methods of pairwise alignment ? Give the one which gives the optimal alignment.
 - (g) Define genetic distance and state two distances based phylogenetic tree prediction methods.
 - (h) What type of information is obtained in a cluster analysis of microarray data ? Explain with a suitable example.
 - (i) What is in-silio secondary structure prediction of protein ?
 - (j) Write note on RASMOL.

SECTION – B

2. Attempt any **three** question parts : **10 × 3 = 30**
- (a) Explain the following : **10**
 - (i) Molecular dynamics in drug designing
 - (ii) Neural networking
 - (b) Explain about the methods employed for sequencing an entire Genome. **10**



- (c) Discuss the ultrametric tree construction. What kind of data is used to build such a tree ? Do the data points 6,3,3 qualify to be ultrametric ? **10**
- (d) How PAM matrix was developed ? How will you choose the best PAM scoring matrix for detecting sequence similarity ? **10**
- (e) Align the following sequence using Smith Waterman Algorithm, Similarity = 1, mismatch = 1, gap = 2, using **10**
- S1 = GCTGGAACCAG
S2 = ACTGGATCAG

SECTION – C

- Attempt **all** questions : **10 × 5 = 50**
3. Attempt any **two** parts : **5 × 2 = 10**
- (a) What is an 'E-Value' ? You do a databank search using FASTA with an amino acid sequence as query. The only reported match has an E-value of 10. What does this mean for the similarity and homology of sequences ?
- (b) Describe DOT-MATRIX sequence alignment.
- (c) Calculate the dynamic programming matrix and optimal alignment for two DNA sequence 'GAATTC' and 'GATTA', scoring '+2' for matching, '-1' for mismatching and '2' for linear gap penalty.
4. Attempt any **one** part : **10 × 1 = 10**
- (a) What is Entaez ? Explain its role in linking different data sources.
- (b) What are interaction databases ? Give examples of them at least two and their features and uses.
5. Attempt any **one** part : **10 × 1 = 10**
- (a) What are the different methods used for DNA sequencing ? Elaborate them with examples.
- (b) Illustrate Edmand Degradation method of protein sequencing.
6. Attempt any **one** part : **10 × 1 = 10**
- (a) What is meant by iterative method of multiple sequence alignment give an algorithm which does this type of alignment ? Explain it and give a tool which performs it.
- (b) Elaborate the applications of bioinformatics on drug discovery and drug design.
7. Attempt any **two** parts : **5 × 2 = 10**
- (a) What are the three methods of gene prediction and give example of program in any one method.
- (b) State two distances based phylogenetic tree prediction methods.
- (c) Compare Oigomer array and cDNA array construction.