Name : .....

Roll No. : .....

Invigilator's Signature : .....

CS/M.Tech (BT)/SEM-2/MBT-201/2011

2011

ADVANCED BIOINFORMATICS

Time Allotted : 3 Hours *Full Marks :* 70

The figures in the margin indicate full marks.

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

GROUP – A

(Multiple Choice Type Questions)

1. Choose the correct alternatives for any *ten* of the following : 10 × 1 = 10

   i) Most micro-arrays consist of a solid support on which is immobilized

   a) DNA b) RNA

   c) Genes d) Transcripts.

   ii) The PAM250 matrix is defined as having an evolutionary divergence in which what percentage of amino acids between two homologous sequences have changed over time ?

   a) 1% b) 20%

   c) 80% d) 250%.

iii) According to molecular clock hypothesis

a) all proteins evolve at the same, constant rate

b) all proteins evolve at a rate that matches the fossil record

c) for every given protein, the rate of molecular evolution gradually slows down like a clock that runs down

d) for every given protein, the rate of molecular evolution is approximately constant in all evolutionary lineages.

iv) You have a protein sequence, and you want to quickly predict its structure. After performing BLAST and PSI-BLAST searches, you identify the most closely related proteins with a known structure as several having 17% amino acid identity to your protein. Which of these options is best ?

a) *X*-ray crystallography

b) NMR

c) Submitting your sequence to a protein structure prediction server that performs homology modelling

d) Submitting your sequence to a protein structure prediction server that performs *ab initio* modelling.

v) An advantage of *X*-ray crystallography relative to NMR for structure determination is that using *X*-ray crystallography

a) it is easier to solve the structure of transmembrane domain-containing proteins

b) it is easier to grow crystals than to prepare samples for NMR

c) it is easier to interpret diffraction data

d) it is easier to determine the structures of large proteins.

vi) You have two distantly related proteins. Which BLOSUM or PAM matrix is best to use to compare then ?

a) BLOSUM45 or PAM250

b) BLOSUM45 or PAM1

c) BLOSUM80 or PAM250

d) BLOSUM80 or PAM1.

vii) If you want literature information, what is the best website to visit ?

a) OMIM b) Entrez

c) Pubmed d) PROSITE.

viii) Which of the following BLAST programs is best used for the analysis of immunoglobulins ?

a) RPS-BLAST

b) PHI-BLAST

c) IgBLAST

d) ProDom.

ix) RNA samples are commonly converted to *c*DNA or *c*RNA for micro-array studies and visualized by labelling with

a) radioactivity or phosphorescence

b) radioactivity or fluorescence

c) radioactivity or RNA probes

d) radioactivity or DNA probes.

x) A major advantage of two-dimensional protein gels as a high throughput technology for protein analysis is that

a) sample preparation and the process or running two dimensional gels is straight forward and can be automated

b) the result of two-dimensional gels includes data on both the size and the charge of thousands of proteins

c) the technique is well suited to the detection of low-abundance proteins

d) the technique is well suited to the detection of hydrophobic proteins.

xi) Which of the following statements best illustrates the theory behind the hidden Markov model (HMM) ?

a) It relies on first creating a phylogenetic tree

b) It calculates the probability of an amino acid occurrence at each position

c) It calculates a multiple sequence alignment based on scores from randomly generated sequences

d) It only aligns sequences that belong to an already described protein family.

xii) You have 200 viral DNA sequences of 500 residues each, and you want to know if there are any pairs that are identical (or nearly identical). Which of the following is the most efficient method to use ?

a) BLAST

b) Maximum-likelihood phylogenetic analysis

c) Neighbour-joining phylogenetic analysis

d) Popset.

GROUP – B

(Short Answer Type Questions)

Answer any *three* of the following. 3 × 5 = 15

2. Write down the different technologies that are used in lead identification ? In what manner quantitative structure-activity relationship (QSAR) is helpful in screening prospective drug molecules ? 2 + 3

3. What is the full form of SNP ? SNPs can help in explaining why different people respond differently to the same drug. Explain how it can be accomplished. 1 + 4

4. What is pair-wise sequence alignment ? Calculate the Log Odds score for changes between Phe and Tyr at an evolutionary distance of 250 Pams. 1 + 4

5. What is MSA ? Write down the flowchart of Global Multiple Sequence Alignment. 2 + 3

6. What is the importance of Data Bank ? What are the applications of RNA structure modelling ? 2 + 3

GROUP – C

(Long Answer Type Questions)

Answer any *three* of the following. 3 × 15 = 45

7. Let S1 = GAATTCAGTTA and S2 = GGATCGA.

   a) Build the advanced Dynamic programming table (Using Needleman/Wunsch Algorithm) for the strings.

   b) List all optimal global alignments between S1 and S2.

   c) What is gap ? What are the different types of gaps that can be assigned to sequence analysis algorithm ?

   d) Mention the difference between local alignment and global alignment. 5 + 2 + 2 + 6

8. What is homology modelling ? Write down the steps to be used for homology modelling. What are the difficulties of homology modelling ? How can you solve or minimize these difficulties ? 3 + 6 + 3 + 3

9. Write short notes on any *five* of the following : 5 × 3

   a) Dot matrix

   b) Identity and similarity

   c) BACs

   d) Motif

   e) PSSM

   f) Sequence database

   g) Bootstrap analysis.

10. What are neural networks ? Why do you use neural networks to predict protein structures ? What do you know about secondary structure and tertiary structure of a protein ? Write two publicly available HMM implementation softwares, How are the processes of energy minimization and molecular dynamics useful in the prediction of protein structures ?

2 + 2 + 4 + 2 + 5

11. What are the advantages and disadvantages of microarray experiment ? Write down the different steps of this experiment.

6 + 9