Group - E

8.(a) Define a force field in the context of molecular modelling. What are its key contributors? Write the full form of a functional force field defining all terms and symbols.

(b) How is estimation of affinity generally done during docking experiments? Define scoring function and its important characteristics.

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

9.(a) Explain the steps of structure based drug design (SBDD) using a flow chart .

(b) How is lead improvement assisted by Quantitative Structure Activity Relationships (QSAR)? Illustrate using an example.

(c) Describe Lipinski's "rule of five" using an example. What is its utility in computational drug design?

(d) Outline the steps of the Metropolis Monte Carlo algorithm.

3 + 3 + 3 + 3 = 12

M.TECH/BT/2ND SEM/BIOT 5201/2017
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 alternative 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) The alignment procedure that tries to align an entire sequence is
(a) multiple sequence alignment (b) local alignment
(c) global alignment (d) proteomic alignment.

(iii) For a good drug likeliness, the number of hydrogen bond donor groups should be
(a) > 7 (b) > 5 (c) > 4 (d) > 15.

(iv) Phylogenetic relationship can be shown by
(a) Dendrogram (b) GENBANK ,
(c) data retrieving tool (d) data search tool.

(v) Following are tools that use sequence alignment at some point except
(a) Rasmol (b) BLAST
(c) FASTA (d) CLUSTALW.

(vi) In a relational database, columns in a table are indexed according to a feature called
(a) a relation (b) a delimiter
(c) an attribute (d) a data record.

(vii) Which of the following DOES NOT result in lead compound generation?
(a) study of drugs effective against similar diseases
(b) study of natural sources
(c) study of metal ions
(d) computer aided screening.

(viii) Pharmacophore generation involves
(a) identification of common structures of many pharmacologically active compounds
(b) identification of chromophores
(c) identification of bioactive compounds of dissimilar activity
(d) identification of fluorogenic substrates.

(ix) A coiled coil is a secondary structural element with two or more
(a) interacting alpha-helices (b) beta-sheets
(c) knots (d) all of the above.

(x) In QSAR, the constant σ is the
(a) the Hammett electronic substituent constant
(b) the lipophilicity consant
(c) the steric clash constant
(d) the Hantszch constant.

Group - B

2.(a) Itemize the main applications of pairwise alignment.

(b) What are the types of algorithms that perform local and global alignment? Outline their differences.

(c) Define a scoring matrix in the context of alignment. Illustrate your answer with two examples.

(d) Explain the reasons for choosing BLOSUM45 as a substitution matrix over BLOSUM80 while using BLAST.

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

3.(a) Compare between heuristic algorithm and exhaustive algorithm.

(b) Mention the basic principles followed by Clustal.

(c) Mention the role of scoring matrix in Clustal in assessment of evolutionary distances.

(d) “Progressive alignment method is not suitable for sequences of uneven sizes” - justify the statement with suitable reasons.

3 + 4 + 2 + 3 = 12

Group - C

4.(a) What portion of a protein is more conserved than any other portion? Describe how position specific scoring matrices (PSSM) are constructed. What is the application of a PSSM?

(b) List the various mechanisms of matching regular expressions with a query.

(2 + 6 + 2) + 2 = 12

5.(a) “Hidden Markov Model (HMM) is a combination of two or more Markov chains” - explain the role of various states and statistical models (probability values).

(b) "In HMM construction the problem of over fitting is often overcome by regularization" - justify the statement.

(c) "Fuzzy matches are also referred to as approximate matches" - justify the statement.

6 + 3 + 3 = 12

Group - D

6.(a) Outline and explain the steps involved in the homology modelling and fold recognition based methods of protein tertiary structure prediction.

(b) Briefly and *critically analyse* the percentage accuracies of each of the above two tertiary structure prediction approaches.

(c) Name the types of RNA secondary structures. Why is tertiary structure prediction for RNA difficult? Use a graph to explain your answer.

5 + 4 + (1 + 2) = 12

7.(a) Define CASP. As of 2016, name and briefly explain the main categories for CASP prediction.

(b) What are artificial neural networks (ANN)? Use *appropriate network diagram(s)* to explain the operation of an ANN including its *hidden* layers. Why is a sigmoidal function preferred over a sharp threshold function for a neuron’s output?

(c) Use a properly labelled diagram ONLY to depict how a neural network has been applied to protein secondary structure prediction.

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