M.TECH/BT/2ND SEM/BIOT 5201/2015
2015

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 x 1=10**

(i) 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.

(ii) The 'training' process in a HMM involves
(a) calculation of the ordering of the residues in each column of the multiple alignment
(b) deconstruction of a PSSM
(c) calculation of the residues in each column of the alignment
(d) all of the above.

(iii) If a overall candidate compound library size is 10^{12} compounds, approximately how many compounds are typically forwarded for consideration for clinical trials?
(a) 1-10 (b) 10-100 (c) 10^3 (d) 10^4.

(iv) Chou Fasman method is based on
(a) neural network method (b) rule based method
(c) information theoretical method (d) none of these.

(v) The side chains of the twenty amino acids vary in
(a) polarity (b) size
(c) shape & rigidity (d) all of these.

(vi) The meaning of the E-value in BLAST is-
(a) The probability that the query sequence and the subject sequence come from the same organism.
(b) The probability that the query sequence and the subject sequence are homologous.
(c) The expected number of generated sequences that would have the observed alignment (or better).
(d) The inverse of the similarity between the query sequence and the subject sequence.

(vii) Rotation around the N-C_α and C_α-C single bonds of a protein backbone is permitted to all amino acids except

(a) tryptophan (b) histidine (c)tyrosine (d) proline.

(viii) Targeting metabolic pathways specific to microorganism, for the purpose of developing drugs,

(a) is more likely to interact with a human homolog
(b) is less likely to interact with a human homolog
(c) has no relevance to interaction with a human homolog
(d) could be (a) or (b) depending on the chemical compound.

(ix) Maximum Likelihood method is used in molecular phylogeny when:

(a) there is maximum similarity among the aligned sequences
(b) there is minimum similarity among the aligned sequences
(c) there is minimum number of hits found in BLAST
(d) all the sequences are from same phylum.
(e)

(x) Pharmacophore is the

(a) lead compound that has the ability to exert a desired biological effect
(b) 3-D arrangement of the compound with the biological molecule that enables a desired biological effect
(c) 3-D arrangement of functional groups necessary for a molecule to exert its effect in a biological syste
(d) None of the above.

Group – B

2. (a) Outline the relationships between homology, similarity and identity.

(b) What is meant by scoring matrix? When would you choose to use BLOSUM45 and BLOSUM80 in BLAST and why?

(c) Enumerate the steps of Dynamic Programming. Using this algorithm, find out the optimal alignment and score with the following 2 sequences.(Write your answer in a step wise manner)

Seq 1: GAATTCAGTTA
Seq 2: GGATCGA **3+(2+2)+5=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 *ab initio* method is said to be not suitable for eukaryotic gene prediction?

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

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

Group – C

4. Define cladogram. The following distances are found among four sequences M, N, O and P: construct the phylogenetic tree with any UPGMA method citing all the steps to regenerate the new distance matrix. Cite the modifications needed on this method mentioning the name of the suitable method over this.

	M	N	O
N	0.40		
O	0.35	0.45	
P	0.60	0.70	0.55

(2+8+1+1)=12

5.(a) Mention in brief the progressive and iterative alignment method along with drawbacks of progressive alignment method.

(b) One of the applications of multiple sequence alignments is identification of related sequences from databases by construction of PSSMs and HMMs- describe briefly about these.

(4+2)+6=12

Group – D

6.(a) What are the three computational approaches to protein 3D structural modelling and prediction? Explain their differences with representative examples.

(b) What are the types of secondary structures for RNA? Why is tertiary structure prediction for RNA difficult? Use an example to highlight how a dot plot is used for RNA structure prediction.

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

7.(a) Why is homology modelling so named? What are the prerequisites of this method? In which case the method of fold recognition is used? Briefly mention the method of fold recognition.

(b) What is the significance of the name of CATH database? Explain the significance of each letter in the name.

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

Group – E

8.(a) The log $(1/K_i)$ of two substituted phenyl-based inhibitors was determined and expected to be a simple linear function of hydrophobicity: log $(1/K_i) = a\pi + c$. Use the data below to develop the corresponding QSAR equation. Explain your answer.

Substituent	Log $(1/K_i)$	π
n-butyl	8.24	2.52
F	7.06	0.63

(b) What are the two conditions for a good lead compound and why? How can molecular modelling and its principles be applied to drug design?

6+(2+4)=12

9.(a) What is meant by docking? How is rigid docking different from flexible docking? Mention the situation when rigid docking would be more applicable and vice versa.

(b) Distinguish between structure-based drug designing and ligand-based drug designing. Mention the importance of ADME parameters in drug designing.

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