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) Which of the following statements is true with respect to the β-barrel membrane protein?
(a) β-strands from the transmembrane segment
(b) The β-strands are ampiphilic
(c) The β-strands contain 5-7 residues
(d) The algorithm for structure prediction for such proteins are non neural network based

(ii) Which of the following represents a data retrieval based bioinformatics tool?
(a) Entrez (b) EMBL
(c) PHD (d) All of these

(iii) Homologation in drug design refers to the
(a) Effect of carbon chain length on drug potency
(b) Effect of carbon chain length on drug toxicity
(c) Effect of aromatic groups on drug potency
(d) Effect of aromatic groups on drug toxicity

(iv) Clustal ω does which of these alignment procedures?
(a) Local alignment (b) Global alignment
(c) Partial alignment (d) Multiple sequence alignment

(v) Which of the following types of algorithms are used in threading based protein tertiary structure prediction?
(a) Pairwise energy based
(b) Profile based
(c) Fitting sequence to a structure database
(d) All of the above

(vi) Which of the following are pairs of orthologues?
(a) Human haemoglobin α and human haemoglobin β
(b) Human haemoglobin α and horse haemoglobin α
(c) Human haemoglobin α and horse haemoglobin β
(d) Human haemoglobin α and human human haemoglobin ϒ

(vii) Which of the following is a wrong statement regarding Gene Prediction Using Markov Models and Hidden Markov Models?
(a) Markov models and HMMs can be very helpful in providing finer statistical description of a gene
(b) A Markov model describes the probability of the distribution of nucleotides in a DNA sequence
(c) In a Markov model the conditional probability of a particular sequence position depends on k alternate positions
(d) A zero-order Markov model assumes each base occurs independently with a given probability

(viii) Pharmacophore generation/identification involves
(a) Identification of common structures of many pharmacologically active compounds
(b) Identification of chromophores
(c) Identification of fluorophores
(d) Identification of bioactive compounds of dissimilar therapeutic activity

(ix) Which of the following is incorrect regarding the advantages of Molecular data for phylogenetics study?
(a) They are more numerous than fossil records
(b) They are easier to obtain as compared to fossil records
(c) Sampling bias is involved
(d) More clear-cut and robust phylogenetic trees can be constructed with the molecular data

(x) In the equation $MR=[(n^2-1)/(n^2+2)$ X $(MW)/d$ for molar refractivity, n represents
(a) density of the compound
(b) thickness of the medium
(c) refractive index of the sodium D line
(d) refractive index.

Group – B

2. (a) Mention the application of sequence alignment.

(b) Name the program in sequence analysis which is based on finding high-scoring ungapped segments among related sequences based on the query sequence. Briefly describe steps of the said procedure which the said programme follows.

(c) Mention the name of the statistical indicator in the result of the above mentioned programme Mention how it is related to raw alignment score.

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

3. (a) What are the three types of algorithms that perform both local and global alignment? Given a particular query sequence, name the sequence analysis program that is based on *finding high scoring ungapped segments* among related sequences. Outline the steps of the procedure followed by this program.

   (b) What is the condition for a statistically significant sequence alignment? Cite the name of the statistical indicator used in the program in part (a) of this question. How is this statistical indicator related to the raw alignment score? Discuss with examples.

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

Group – C

4. (a) Define Markov chain, zero order and first order Markov model cite relationships among zeroorder model and first order model

   (b) Briefly describe role of the following factors in context of Hidden Markov Model. - observed and nonobserved factors

   (c) Graphically represent the three states in a Hidden Markov Model.

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

5. (a) Define a position specific scoring matrix (PSSM). What are its essential characteristics? Explain stepwise how a PSSM may be constructed from a MSA of nucleotide sequences.

   (b) Construct the PSSM for the following sequence set :

Position	1 234 5 6
Sequence 1	ATGTCG
Sequence 2	AAGACT
Sequence 3	TACTCA
Sequence 4	CCGAGG
Sequence 5	AACCTG

(2 + 4) + 6 = 12

Group – D

6. (a) Define CASP in the context of protein tertiary structure prediction. What were the three traditional categories in CASP for protein structure prediction? Explain briefly the nature of the target in each category. Cite four new specialized categories of prediction challenge in CASP. Briefly describe two most recent methodological improvements in CASP data analysis.

   (b) Use a schematic to explain the logic of artificial neural networks with a 3 point input. On what principles were these computational structures originally based on? Use a detailed schematic diagram to represent the application of a 3 layer neural network algorithm to a secondary structure prediction

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

7. (a) What are the important cellular functions and biomedical applications of transmembrane (TM) proteins? Use a schematic of the positive-inside rule to explain why and how special algorithms are needed to solve the secondary structure of TM proteins. Cite two specific factors that can improve the accuracy of algorithms for predicting the secondary structure of TM proteins.

(b) Comparatively tabulate the *methodological differences* in building a tertiary structure model of a protein using homology modelling *vs.* threading. What are the two specific *quantitative* requirements for successful fold recognition by threading?

(c) What two principles form the basis of *ab-initio* protein structure prediction? Outline and briefly explain the steps of the *ab-initio* protein structure prediction algorithm ROSETTA. What are two current methodological limitations of such ab-initio algorithms?

(2 + 2 + 1) + (2 × 2) + 3 = 12

Group – E

8. (a) In *protein structure comparison by alignment* the similarity score s plays an important role. The similarity s in some aspects bears an inverse relationship to the rmsd value in usage for protein structure comparison by the intermolecular method. Describe the essential steps of the intermolecular method of protein structure comparison. Based on your answer , *interpret t*he following equation for similarity value s, $s= a/ [({}^{A}V_{il} - {}^{B}V_{fm})^2 + b$ for the matrix element (l,n). (Hint: use the relationship of the two vectors V_{il} and V_{fm} to the similarity value s to answer this question).

(b) (i) What are the main applications of quantitative structure activity relationships in computer assisted drug design?

(ii) The log (K_i) of two substituted phenyl based inhibitors were determined and expected to be a simple linear function of hydrophobicity:
$\log (1/K_i) = a\pi + c$. *Use the data provided herein to develop the QSAR equation.*

Substituent	Log (1/Ki)	π
n-isopropyl	7.41	2.32
Cl	6.71	0.28

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

9. (a) The Hantsch multiple regression analysis based equation relating the molar concentration of a potential drug to physico-chemical parameters and steric/size effects is given by
$\text{Log } 1/C = -a\pi^2 + b\pi + \rho\sigma + cE_s + d|S + e.$
Define the terms in this equation.

(b) Aspirin, acetaminophen (paracetamol), ibuprofen are non steroidal anti inflammatory drugs (NSAIDs). What are their specific therapeutic uses? Name the enzymes (two) that these drugs are targeted against. What are the functions of these two enzymes? How can you selectively target one of these enzymes so

as to reduce the possibility of adverse side effects of any one of these drugs? What is the mechanism for the adverse side effect of aspirin?

(c) Docking and subsequent scoring of ligand-target complexes are important steps between a total library and testing a small number of compounds for further validation. Answer these two questions relevant to the above statement (i) Name a docking experiment that aims to understand the specificity of a ligand/potential drug(ii) Draw a simple *representative*"inverted funnel" to show how a large library of potential lead compounds can be reduced to an acceptable number for validation testing.

(d) Methods of bioinformatics can be used to prioritize selection of protein targets for experimental structure determination that provide maximal information payoff. Itemize systematically the goals of target selection with examples wherever feasible.

3 + 3 + 3 + 3 = 12

Department & Section	Submission Link
BT	https://classroom.google.com/c/MzQzMzI3NTYwMDM1/a/Mzc0Mjc0NjAyNTk1/details