BIOINFORMATICS
(BIOT 3102)

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 one of the following choices is NOT representative of a bioinformatics experimental application?
(a) genome comparison
(b) gene expression profiling
(c) protein structure comparison
(d) diffusion coefficient calculations.

(ii) BLAST X program is used for
(a)translating a protein sequence
(b)translating a DNA database
(c)translating an input sequence
(d) none of these.

(iii) Which one of the following bioinformatics secondary structure prediction tools is based on an ab-initio method?
(a) PSIPRED
(b) Chou Fasman
(c) PHD
(d) PORTER.

(iv) Which one of the following steps is NOT a step of the *fold recognition based* pairwise energy method for 3D structure prediction of proteins?
(a)Multiple sequence alignment of folds
(b)model building
(c)energy calculations
(d)building a sequence profile.

(v) PDB is a
(a)primary database for macromolecules
(b)composite database
(c)database for three dimensional structure of biological macromolecules
(d)all of these.

(vi) BLOSUM matrices are used for
(a) Multiple sequence alignment
(b) Pairwise sequence alignment
(c) Phylogenetic analysis
(d)all of the above.

(vii) Protein structure prediction depends upon which one of these choices
(a) sequence alignment data
(b) phylogenetic tree construction
(c) gene expression profiling
(d) all of the above.

(viii) Which one of the choices accounts for easier annotation of prokaryotic genes?
(a) they have low gene density
(b) they have split gene structure
(c) they have high gene density and noninterrupted genes
(d) none of the above.

(ix) Which of these is a data retrieval tool?
(a) EMBL
(b) ENTREZ
(c) PHD
(d) all of these.

(x) SWISSPROT is related to
(a) Portable data
(b) a sequence data bank
(c) Swissbank data
(d) none of the above.

Group – B

2. (a) What are the characteristics of a biological database? Using examples highlight the differences between a relational database management system (RDBMS) and a flat file based one. Name one programming language (PL) that is typically used for relational and object oriented databases respectively.

(b) Name two common information retrieval systems for biological databases. What is one common characteristic of both of these retrieval systems? How can complex search queries be performed in a database?

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

3. (a) "Bio-computational tool development forms the foundation of bioinformatics analysis". Based on this statement, use a diagram to show the three major bioinformatics analysis categories where these tools are applied. Use the examples of protein structure prediction and gene annotation *to show in simple terms* how these categories are interconnected.

   (b) (i) How have tools of bioinformatics been utilized in agricultural biotechnology? Highlight your answer with specific examples.
   (ii) What characteristics of machine learning have helped in designing bioinformatics tools that can infer biological properties from database analysis? Name two numerical methods applied to such database analysis.

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

Group – C

4. (a) Define sequence alignment.

   (b) Enumerate the importance of sequence alignment in biology and bioinformatics.

   (c) What are the different types of gaps that are applied in sequence analysis? Characterize these gaps with suitable examples.

   (d) "The scoring systems in dynamic programming based alignment are referred to collectively as a substitution matrix". Explain this statement and briefly explain stepwise how such a scoring system is computed if the sequence is an amino acid one.

2 + 3 + 3 + 4 = 12

5. (a) Enumerate stepwise the procedure of pairwise sequence alignment using Dynamic Programming. Explain your answer with an example utilizing a 2×2 matrix.

   (b) What are the *special features* of the ab-initio based approaches of gene prediction? How can these *features* be utilized for the prediction/annotation of both prokaryotic and eukaryotic genes?Your answer should be in simple understandable terms.

(4 + 2) + 6 = 12

Group – D

6. (a) What is the full form of PERL? Itemize and explain three distinguishing characteristics of PERL programming language. *Your answer should highlight characteristics that made PERL ideal for sequence analysis tasks in biology*.

   (b) Name the common variables that are used as part of PERL programming. Using suitable examples explain the roles of these variables.

(1 + 5) + 6 = 12

7. Write the code for a PERL program that (i) scores two strings in two variables and (ii) then checks if the two strings given as arguments are reverse complements of

each other. Use the PERL built-in functions split, pop, shift and eq (an operator) and then prints out the output.

4 + (4 × 2) = 12

Group – E

8. (a) SCOP is a protein structure classification database. What is its full form? Use a diagram to show how it organizes the protein structures. In a stepwise fashion, explain the logic behind the hierarchy of structures in this database.

   (b) What are the direct applications of predicting protein secondary structures? What are the reasons that have made a relatively simple computational problem like protein secondary structure prediction (α-helices,β sheets) difficult to reach high prediction accuracy? Name the two types of algorithmic methods used for secondary structure of proteins.

   (c) Draw a labelled schematic diagram for a protein SSE prediction algorithm that uses neural networks. Explain the steps in this prediction algorithm. What are the reasons that make the accuracy of such prediction algorithms higher with use of multiple sequence alignments and neural networks?

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

9. (a) What is the need for computer generated determination of the three dimensional structure of the protein? Once you have the computer aided 3D model of a protein, what uses of biochemical relevance can you put that modelled structure to?

   (b) Name the three computational approaches to protein 3D modelling and prediction. Tabulate their basic methodological characteristics.

   (c) Write the steps of the 3D structure modelling procedure of ANY ONE of the three computational approaches in part (b).

(2 + 2) + 5 + 3 = 12

Department & Section	Submission Link
BT	https://classroom.google.com/c/MjM3MzczODM5NzA0/a/Mjc0ODU0Nzc4NDE0/details