



BRAINWARE UNIVERSITY

Term End Examination 2023
Programme – M.Tech.(CSE)-AIML-2022
Course Name – Bio-Informatics
Course Code - PEC-MCSM201A
(Semester II)

LIBRARY
Brainware University
Barasat, Kolkata -700125

Full Marks : 60

Time : 2:30 Hours

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

Group-A

(Multiple Choice Type Question)

1 x 15=15

1. Choose the correct alternative from the following :

- (i) List the main types of data used in bioinformatics.
- | | |
|---------------------------------|---------------------------------|
| a) Numeric and text data | b) Image and audio data |
| c) Biological and chemical data | d) Sequence and structural data |
- (ii) Select the correct option. The laboratory work using computers and computer-generated models generally offline is referred to as _____.
- | | |
|--------------|---------------------|
| a) In silico | b) Wet lab |
| c) Dry lab | d) All of the above |
- (iii) Identify the correct option. The laboratory work using computers and associated with web-based analysis generally online is referred to as _____.
- | | |
|--------------|---------------------|
| a) In silico | b) Wet lab |
| c) Dry lab | d) All of the above |
- (iv) Identify what KEGG stands for.
- | | |
|---|---|
| a) Korean Encyclopedia of Genetic Gaps | b) Kazakh Encyclopedia of Genomic Geography |
| c) Kongo Encyclopedia of Genetic Groups | d) Kyoto Encyclopedia of Genes and Genomes |
- (v) Select the year in which the SWISSPROT protein sequence database began.
- | | |
|---------|---------|
| a) 1988 | b) 1987 |
| c) 1986 | d) 1985 |
- (vi) Identify why the NCBI database is important for molecular biology research.
- | | |
|--|--|
| a) It offers a platform for sharing and collaborating on molecular biology data | b) It contains information on gene expression and regulation in various organisms |
| c) It allows for the prediction of protein structure based on amino acid sequence data | d) It provides access to a wide range of molecular biology databases, including GenBank, PubMed, and BLAST |

- (vii) Select how you can search for DNA sequences in the GenBank database.
- a) By browsing through the database using filters based on organism or gene function
b) By entering a keyword or accession number in the search bar
c) By uploading a FASTA sequence file for the DNA of interest
d) By accessing the DNA sequence alignment tool within the database
- (viii) Choose which of the following is a DNA sequence.
- a) AUGUUCGCAUAA
b) AGCUUCGGAAUA
c) ACUGUUGCGUAA
d) None of the above
- (ix) Calculate the percentage identity of two aligned sequences and interpret the result. Given two aligned sequences with a total length of 120 amino acids, 90 of which are identical, what is the percentage identity?
- a) 0.3
b) 0.75
c) 0.25
d) 0.9
- (x) Choose the correct mRNA sequence that would be transcribed from the following DNA sequence: TACGCTAGCG
- a) AUGCGAUAGC
b) AUGCGUAUGC
c) UACGCUAGCG
d) None of the above
- (xi) Select the type of homology that can be classified as orthologous.
- a) Homology between a gene and a non-coding DNA sequence
b) Homology between two genes in the same species
c) Homology between two genes in different species
d) Homology between a protein and a nucleotide sequence
- (xii) Determine the amino acid sequence encoded by the following RNA sequence: AUGCCAUGUUGA
- a) Met-Pro-Cys
b) Met-His-Leu
c) Arg-Cys-Asn
d) None of the above
- (xiii) Differentiate between the following types of mutations: point mutation and frame-shift mutation
- a) Point mutation involves insertion or deletion of nucleotides while frame-shift mutation involves substitution of nucleotides
b) Point mutation involves substitution of a single nucleotide while frame-shift mutation involves insertion or deletion of nucleotides
c) Point mutation and frame-shift mutation are the same thing
d) Point mutation involves substitution of nucleotides while frame-shift mutation involves inversion of nucleotides
- (xiv) Estimate the percentage of GC content in the following DNA sequence. Sequence : AGGTTACGCTACGTAGGACTGACTGCGT
- a) 0.5
b) 0.6
c) 0.53
d) 0.46
- (xv) Estimate the GC content of a DNA sequence with 200 base pairs if it contains 70 cytosine (C) nucleotides and 60 guanine (G) nucleotides.
- a) 0.35
b) 0.6
c) 0.65
d) 0.7

Group-B

(Short Answer Type Questions)

2. Define Bioinformatics. (3)
3. Explain a BLAST search. (3)
4. Explain briefly the various types of protein structures. (3)

5. Distinguish between DNA replication and DNA transcription. (3)
6. Explain the betweenness property in a PPI network. (3)

OR

Explain how a peptide bond formed between two amino acids. (3)

Group-C

(Long Answer Type Questions)

5 x 6 = 30

7. Construct a maximum likelihood tree based on amino acid sequences of a specific protein family. Explain the methods used for tree construction and the interpretation of the results. (5)
8. State the significance of pathway databases in bioinformatics. (5)
9. Write a Perl script to extract all protein sequences from a FASTA file and write them to a new file. (5)
10. Explain the difference between identity and similarity searches, and describe how these types of searches are used in bioinformatics research. (5)
11. Explain the basic principles of Perl programming and its applications in bioinformatics research. (5)
12. Write how bioinformatics can be used to personalize cancer treatment based on a patient's genomic profile. (5)

OR

Write about a hidden Markov model (HMM) and how it is used in bioinformatics. (5)
