



BRAINWARE UNIVERSITY

Term End Examination 2020 - 21

Programme – Master of Technology in Computer Science & Engineering

Course Name – Bio-informatics

Course Code - PEC-MCS302A

Semester / Year - Semester III

Time allotted : 85 Minutes

Full Marks : 70

[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 70=70

1. (Answer any Seventy)

(i)

What is the length of a motif, in terms of amino acids residue?

- | | |
|----------|----------|
| a) 30-60 | b) 10-20 |
| c) 70-90 | d) 1-10 |

(ii) Which of the following is false about the 'loop' structure in proteins?

- | | |
|--|--|
| a) They connect helices and sheets | b) They are more tolerant of mutations |
| c) They are more flexible and can adopt multiple conformations | d) They are never the components of active sites |

(iii) Which of the common structural motifs are described wrongly?

- | | |
|--|--|
| a) ?-hairpin – adjacent antiparallel strands | b) Greek key – 4 adjacent antiparallel strand |
| c) ?-?-? – 2 parallel strands connected by helix | d) ?-?-? – 2 antiparallel strands connected by helix |

(iv) Motifs that can form ?? horseshoes conformation are rich with which protein residue?

- | | |
|------------|-------------|
| a) Proline | b) Arginine |
| c) Valine | d) Leucine |

(v) The protein structural motif domain- helix loop helix are contained by all of the following except _____

- a) Scleraxis
- b) Neurogenins
- c) Transcription Factor 4
- d) Leucine zipper

(vi) In the zinc finger, which residues in this sequence motif form ligands to a zinc ion?

- a) Cysteine and histidine
- b) Cysteine and arginine
- c) Histidine and proline
- d) Histidine and arginine

(vii) Which of the following is not a member database of InterPro?

- a) SCOP
- b) HAMAP
- c) PANTHER
- d) Pfam

(viii) What is the source of protein structures in SCOP and CATH?

- a) Uniprot
- b) Protein Data Bank
- c) Ensemble
- d) InterPro

(ix) Which of the following statements about SUPERFAMILY database is incorrect regarding its features?

- a) Sequences can be submitted raw or FASTA format
- b) Sequences must be submitted in FASTA format only
- c) It searches the database using a superfamily, family, or species name plus a sequence, SCOP, PDB or HMM ID's
- d) It has generated GO annotations for evolutionarily closed domains and distant domains

(x) When did Needleman-Wunsch first describe the algorithm for global alignment?

- a) 1899
- b) 1970
- c) 1930
- d) 1950

(xi) Which of the following is not an advantage of Needleman-Wunsch

algorithm?

- a) New algorithmic improvements as well as increasing computer capacity make possible to align a query sequence against a large DB in a few minutes
- b) Similar sequence region is of same order and orientation
- c) This does not help in determining evolutionary relationship
- d) If you have 2 genes that are already understood as closely related, then this type of algorithm can be used to understand them in further details

(xii) Which of the following does not describe global alignment algorithm?

- a) In initialization step, the first row and first column are subject to gap penalty
- b) Score can be negative
- c) In trace back step, beginning is with the cell at the lower right of the matrix and it ends at top left cell
- d) First row and first column are set to zero

(xiii) Which of the following is untrue regarding the scoring system used in dynamic programming?

- a) If the residues are same in both the sequences the match score is assumed as +5 which is added to the diagonally positioned cell of the current cell
- b) If the residues are not same, the mismatch score is assumed as -3
- c) If the residues are not same, the mismatch score is assumed as 3
- d) The score should be added to the diagonally positioned cell of the current cell

(xiv) When did Smith–Waterman first describe the algorithm for local alignment?

- a) 1950
- b) 1970
- c) 1981
- d) 1925

(xv) Local alignments are more used when _____

- a) There are totally similar and equal length
- b) Dissimilar sequences are suspected to

sequences

contain regions of similarity

c) Similar sequence motif with larger sequence context

d) Partially similar, different length and conserved region containing sequences

(xvi) Which of the following is untrue regarding the gap penalty used in dynamic programming?

a) Gap penalty is subtracted for each gap that has been introduced

b) Gap penalty is added for each gap that has been introduced

c) The gap score defines a penalty given to alignment when we have insertion or deletion

d) Gap open and gap extension has been introduced when there are continuous gaps (five or more)

(xvii) Which of the following does not describe k-tuple methods?

a) k-tuple methods are best known for their implementation in the database search tools FASTA and the BLAST family

b) They are also known as words methods

c) They are basically heuristic methods to find local alignment

d) They are useful in small scale databases

(xviii) Which of the following is untrue regarding BLAST and FASTA?

a) FASTA is faster than BLAST

b) FASTA is the most accurate

c) BLAST has limited choices of databases

d) FASTA is more sensitive for DNA-DNA comparisons

(xix) Which of the following scores are not considered while calculating the SP scores?

a) All possible pair wise matches

b) All possible mismatches

c) All possible gap costs

d) Number of gap penalties

(xx) Which of the following is untrue regarding Progressive Alignment Method?

a) Progressive alignment depends on the stepwise assembly of multiple alignment

b) It speeds up the alignment of multiple sequences through a multistep process

and is heuristic in nature

- c) It first conducts pair wise alignments for each possible pair of sequences using the Needleman–Wunsch global alignment method and records these similarity scores from the pair wise comparisons
- d) Its drawback is it slows down the alignment of multiple sequences through a single step process

(xxi) Which of the following is not a drawback of the progressive alignment method?

- a) The progressive alignment method is not suitable for comparing sequences of different lengths because it is a global alignment–based method.
- b) In this method the use of affine gap penalties, long gaps are not allowed, and, in some cases, this may limit the accuracy of the method
- c) In this method the use of affine gap penalties, long gaps is allowed, and, in some cases, this may limit the accuracy of the method
- d) The final alignment result is also influenced by the order of sequence addition

(xxii) Which of the following is untrue about iterative approach?

- a) The iterative approach is based on the idea that an optimal solution can be found by repeatedly modifying existing suboptimal solutions
- b) Because the order of the sequences used for alignment is different in each iteration
- c) This method is also heuristic in nature and does not have guarantees for finding the optimal alignment
- d) This method is not based on heuristic methods

(xxiii) Which of the following is untrue about DIALIGN2?

- a) It is a web based program designed to detect local similarities
- b) It is designed to detect global similarities
- c) It does not apply gap penalties and thus is not sensitive to long gaps
- d) The method breaks each of the sequences down to smaller segments and performs all possible pair wise alignments between the segments

(xxiv) Which of the following is untrue about Protein substitution matrices?

- a) They are significantly more complex than DNA scoring matrices
- b) They have the $N \times N$ matrices of the amino acids
- c) Protein substitution matrices have quite important role in evolutionary studies
- d) They are significantly quite less complex than DNA scoring matrices

(xxv) Which of the following is not one of the requirements for implementing algorithms for sequence database searching?

- a) Size of the dataset
- b) Sensitivity
- c) Specificity
- d) Speed

(xxvi) Which of the following is incorrect?

- a) Smith–Waterman algorithm is the fastest
- b) Smith–Waterman algorithm is comparatively slower method
- c) To speedup up comparison, heuristic methods are used
- d) Heuristic algorithms perform faster searches

(xxvii) In sequence alignment by BLAST, each word from query sequence is typically _____ residues for protein sequences and _____ residues for DNA sequences.

- a) ten, eleven
- b) three, three
- c) three, eleven
- d) three, ten

(xxviii) Which of the following is not a correct about FASTA?

- a) Its stands for FAST ALL
- b) It was in fact the first database similarity search tool developed, preceding the development of BLAST
- c) FASTA uses a ‘hashing’ strategy to find matches for a short stretch of identical residues with a length of k
- d) The string of residues is known as blocks

(xxix) The positional difference for each word between the two sequences is

obtained by _____ the position of the _____ sequence from that of the _____ sequence and is expressed as the offset.

- a) subtracting, second, first
- b) adding, second, first
- c) adding, first, second
- d) subtracting, first, second

(xxx) _____ the smallest amino acid, has a hydrogen atom as the R group.

- a) valine
- b) proline
- c) Glycine
- d) threonine

(xxxii) Which of the following is not correct?

- a) The rigid double bond structure forces atoms associated with the peptide bond to lie in the same plane, called the dipeptide plane
- b) A peptide bond is actually a partial double bond owing to shared electrons between O=C–N atoms
- c) Because of the planar nature of the peptide bond and the size of the R groups, there are considerable restrictions on the rotational freedom by the two bonded pairs of atoms around the peptide bond
- d) The angle of rotation about the bond is referred to as the dihedral angle (also called the torsional angle)

(xxxiii) Which of the following is not correct about the Coils and Loops?

- a) They are regular structures
- b) They are irregular structures
- c) The loops are often characterized by sharp turns or hairpin-like structures
- d) If the connecting regions are completely irregular, they belong to random coils

(xxxiiii) Which of the following is untrue about SCOP?

- a) It is a database for comparing and classifying protein structures
- b) It is constructed almost entirely based on manual examination of protein structures
- c) The proteins are grouped into hierarchies of classes, folds, superfamilies, and families having low sequence identity (>30%)
- d) The SCOP families consist of proteins

(xxxv) In CATH, Structural domain separation is carried by

- a) manual comparison only
- b) computer programs only
- c) human expertise only
- d) a combined effort of a human expert and computer programs

(xxxv) Which of the following is untrue regarding the drawbacks of expressed sequence tags (ESTs)?

- a) They are often of low quality because they are automatically generated without verification
- b) Many bases are ambiguously determined, represented by N's
- c) Frame shift errors and artifactual stop codons are some common errors
- d) Despite of all the failures, the translation the sequences is smooth

(xxxvi) Which of the following is untrue regarding UniGene?

- a) It is an NCBI EST cluster database.
- b) Overlapping EST sequences are computationally processed to represent a single expressed gene.
- c) Each cluster is a set of overlapping EST sequences
- d) The overlapping EST sequences are computationally processed to represent a set of expressed genes

(xxxvii) Which of the following is untrue regarding SAGE?

- a) It stands for Serial analysis of gene expression
- b) It is another high throughput, sequence-based approach for global gene expression profile analysis
- c) It stands for Squared analysis of gene expression
- d) Unlike EST sampling, SAGE is more quantitative in determining mRNA expression in a cell

(xxxviii) Which of the following is incorrect about Oligonucleotide Design in A microarray?

- a) DNA microarrays are generated by fixing oligonucleotides onto a solid support
- b) The oligonucleotide array slide represents thousands of preselected genes from an organism

c) The length of oligonucleotides is typically in the range of twenty-five to seventy bases long

d) The oligonucleotides don't react with cDNA samples

(xxxix) Which of the following is incorrect about Classification of microarray data?

a) For microarray data, clustering analysis identifies coexpressed and coregulated genes

b) For microarray data, clustering analysis identifies coexpressed but not coregulated genes

c) For microarray data, clustering analysis identifies and coregulated but not coexpressed genes

d) Genes within a category have more similarity in expression than genes from different categories

(xl) Which of the following is incorrect about k-Means Clustering?

a) k-means clustering produces a dendrogram

b) It classifies data through a single step partition

c) It is a divisive approach

d) In this method, data are partitioned into k-clusters, which are prespecified at the outset

(xli) Which of the following is incorrect about Self-Organizing Maps?

a) Clustering by SOMs is in principle similar to the k-means method

b) It doesn't involve neural networks

c) The data points are initially assigned to the nodes at random

d) It starts by defining a number of nodes

(xlii) If the bait and prey proteins _____ they bring the DNA-binding and trans-activation domains in such close proximity that they reconstitute the function of the transcription activator, turning _____ the expression of a reporter gene as a result. Which of the following is not the correct pair of blanks?

a) physically interact, on

b) do not interact, on

c) do not interact, off

d) stop interacting, off

(xliii) The justification behind Rosetta stone method is that when two domains are fused in a single protein, they have to be in _____ proximity to perform a common function.

- | | |
|----------------------|--------------------|
| a) distant | b) close |
| c) extremely distant | d) extremely close |

(xliv)

Related sequences are identified through the database similarity searching and as the process generates multiple matching sequence pairs, it is often necessary to convert the numerous pair wise alignments into a single alignment.

- | | |
|---------------|------------|
| a) | b) |
| True | False |
| c) | d) |
| None of these | inadequate |

(xlv)

The scoring function for multiple sequence alignment is based on the concept of sum of pairs (SP).

- | | |
|---------------|------------|
| a) | b) |
| True | False |
| c) | d) |
| None of these | inadequate |

(xlvi)

There are two approaches viz. exhaustive and heuristic approaches used in multiple sequence alignment.

- | | |
|------|-------|
| a) | b) |
| True | False |

c)

None of these

d)

inadequate

(xlvii)

As the amount of computational time and memory space required increases exponentially with the number of sequences, it makes the multidimensional search matrix method computationally prohibitive to use for a large data set.

a)

True

c)

None of these

b)

False

d)

inadequate

(xlviii)

The major drawback of the progressive and iterative alignment strategies is that they are largely global alignment based and may therefore fail to recognize conserved domains and motifs among highly divergent sequences of varying lengths.

a)

True

c)

None of these

b)

False

d)

inadequate

(xlix)

The global sequence alignment is suitable when the two sequences are of dissimilar length, with a negligible degree of similarity throughout.

a)

True

b)

False

c)

d)

None of these

inadequate

(l)

The substitution matrices are rarely used in this type of matching.

a)

b)

True

False

c)

d)

None of these

inadequate

(li)

In Needleman-Wunsch algorithm, the gaps are scored -2.

a)

b)

True

False

c)

d)

None of these

inadequate

(lii)

Sensitivity refers to the ability to find as many correct hits as possible.

a)

b)

True

False

c)

d)

None of these

inadequate

(lii)

In heuristic methods, speed doesn't vary with the size of database.

- | | |
|---------------|------------|
| a) | b) |
| True | False |
| c) | d) |
| None of these | inadequate |

(liv)

In sequence alignment by BLAST, the second step is to search a sequence database for the occurrence of these words.

- | | |
|---------------|------------|
| a) | b) |
| True | False |
| c) | d) |
| None of these | inadequate |

(lv)

If one is looking for protein homologs encoded in newly sequenced genomes, one may use TBLASTN, which translates nucleotide database sequences in all six open reading frames.

- | | |
|---------------|------------|
| a) | b) |
| True | False |
| c) | d) |
| None of these | inadequate |

(lvi)

The first step in FASTA alignment is to arrange the sequences in matrices' rows and columns in

order to be analyzed.

a)

b)

True

False

c)

d)

None of these

inadequate

(lvii)

The second step in FASTA is to narrow down the high similarity regions between the two sequences.

a)

b)

True

False

c)

d)

None of these

inadequate

(lviii)

FASTX, which compares a protein query sequence to a translated DNA database.

a)

b)

True

False

c)

d)

None of these

inadequate

(lix)

Z-score describes the number of standard deviations from the mean score for the database search.

a)

b)

True

False

c)

d)

None of these

inadequate

(lx)

Within the hydrophobic set of amino acids, they can be further divided into aliphatic and aromatic.

a)

b)

True

False

c)

d)

None of these

inadequate

(lxi)

The peptide formation involves two amino acids covalently joined together between the carboxyl group of one amino acid and the amino group of another.

a)

b)

True

False

c)

d)

None of these

inadequate

(lxii)

The classification results from both systems, SCOP and CATH are quite dissimilar.

a)

b)

True

False

c)

d)

None of these

inadequate

(lxiii)

The second step in structure classification is to separate structurally distinct domains within a structure.

a)

b)

True

False

c)

d)

None of these

inadequate

(lxiv)

The last step in structure classification involves grouping proteins/domains of similar structures.

a)

b)

True

False

c)

d)

None of these

inadequate

(lxv)

To generate EST data, clones in the cDNA library are randomly selected for sequencing from either end of the inserts.

a)

b)

True

False

c)

d)

None of these

inadequate

(lxvi)

GenBank has a special EST database, dbEST that contains EST collections for a large number of organisms.

- | | |
|---------------|------------|
| a) | b) |
| True | False |
| c) | d) |
| None of these | inadequate |

(lxvii)

A supervised analysis refers to classification of data into a set of predefined categories. For example, depending on the purpose of the experiment, the data can be classified into predefined 'diseased' or 'normal' categories

- | | |
|---------------|------------|
| a) | b) |
| True | False |
| c) | d) |
| None of these | inadequate |

(lxviii)

TIGR TM4 is a suite of multiplatform programs for analyzing microarray data.

- | | |
|---------------|------------|
| a) | b) |
| True | False |
| c) | d) |
| None of these | inadequate |

(lxix)

An alternative approach to determining protein–protein interactions is to use a large-scale affinity purification technique that involves attaching fusion tags to proteins and purifying the associated protein complexes in an affinity chromatography column.

a)

True

c)

None of these

b)

False

d)

inadequate

(lxx)

When the two domains are located in two different proteins, to preserve the same functionality, their close proximity and interaction have to be preserved as well.

a)

True

c)

None of these

b)

False

d)

inadequate