

Name : .....

Roll No. : .....

Invigilator's Signature : .....

**CS/B.Tech(BT)/SEM-5/BT-503/2009-10  
2009**

**BIOINFORMATICS – I**

Time Allotted : 3 Hours

Full Marks : 70

*The figures in the margin indicate full marks.*

*Candidates are required to give their answers in their own words as far as practicable.*

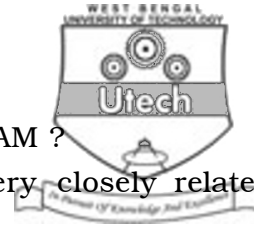
**GROUP – A**

**( Multiple Choice Type Questions )**

1. Choose the correct alternatives for any *ten* of the following :

10 × 1 = 10

- i) The first bioinformatics database was created by
  - a) Richard Durbin
  - b) Dayhoff
  - c) Steven Altschul
  - d) David Lipman.
- ii) Which of the following database can be used to access protein domain information ?
  - a) Prosite
  - b) DDBJ
  - c) SANGER
  - d) KEGG.
- iii) If you want literature information, what is the best website to visit ?
  - a) OMIM
  - b) Entrez
  - c) PubMed
  - d) PROSITE.



- iv) How does the BLOSUM differ from PAM ?
- a) It is best used for aligning very closely related proteins
  - b) It is based on local multiple alignments from distantly related proteins
  - c) It is based on global multiple alignment from closely related proteins
  - d) It combines local and global alignment information.
- v) Which of the following compares a protein query sequence against a translated nucleotide sequence library ?
- a) FASTA
  - b) FASTF
  - c) TFASTA
  - d) FASTX.
- vi) As the  $E$  value of a BLAST search becomes smaller
- a)  $K$  value also becomes smaller
  - b) Score tends to be larger
  - c) Probability  $p$  tends to be larger
  - d) The extreme value distribution becomes less skewed.
- vii) The main difference between Pfam-A and Pfam-B is that
- a) Pfam-A is manually curated while Pfam-B is automatically curated.
  - b) Pfam-A uses HMMs while Pfam-B does not.
  - c) Pfam-A provides full length protein alignments while Pfam-B aligns protein fragments.
  - d) Pfam-A incorporates data from SMART and PROSITE while Pfam-B does not.
- viii) In a position-specific scoring matrix, the score for any given amino acid residue is assigned based on
- a) a PAM or BLOSUM matrix
  - b) its background frequency of occurrence
  - c) the score of its neighbouring amino acid
  - d) its frequency of occurrence in an MSA.



- ix) The BLAST algorithm compiles a list of word. Words at or above a threshold value T are defined as
- Hits and are used to scan the database for exact matches that may then be extended.
  - Hits and are used to scan a database for exact or partial matches that may then be extended
  - Hits and are aligned to each other
  - Hits and are reported as raw score.
- x) You have two distantly related proteins. Which BLOSUM or PAM matrix is best to use to compare them ?
- BLOSUM45 or PAM250
  - BLOSUM45 or PAM1
  - BLOSUM80 or PAM250
  - BLOSUM80 or PAM1.
- xi) Which of the following is not a scalar variable ?
- \$DNA
  - \$23\_Seq
  - \$seq\_dna
  - \$DNA23.
- xii) Let @ base = ( 'A', 'T', 'G', 'C' ) and \$nt = 'X'. The syntax splice @base.1.1.\$nt ) will generate the output
- ATXGC
  - AXTGC
  - AXGC
  - A\$ntTGC.

### GROUP – B

#### ( Short Answer Type Questions )

Answer any *three* of the following.  $3 \times 5 = 15$

- What is Bioinformatics ? Explain its application in genomics and proteomics. 5
- Describe progressive alignment for MSA. 5
- Describe in words how BLAST algorithm works and explain different BLAST programs. 5
- Write a program that takes as input the ID of a gene as well as its sequence and thereafter generates the following output :  
 ID : BC12345  
 Sequence : ATTACATTACA  
 Reverse sequence : ACATTACATTA  
 No of nucleotide : 11. 5
- Explain Motif. Discuss the line given below :  
 [ GA ] – [ IMFAT ] - H - { S } - X ( 3,5 ) - [ GP ]. 5



**GROUP – C**

**( Long Answer Type Questions )**

Answer any *three* of the following.  $3 \times 15 = 45$

7. a) Differentiate local and global alignment. 6  
b) Compute the global alignment ( proposed by Needleman and Wunsch-1970 ) between the two sequences S1 = ABCNJRQCLCRPM and S2 = AJCJNRCKCRBP. 7  
c) Is the alignment you found unique, or are there multiple alignments ? 2
8. What is Scoring or Substitute matrix ? Describe BLOSUM in detail with an example.  $7 \frac{1}{2} + 7 \frac{1}{2}$
9. a) Why do you use Perl ? What arguments do you frequently use for the Perl interpreter and what do they mean ? 3  
b) What do the symbols \$ @ and % mean when prefixing a variable ? 3  
c) Write a program that takes as input a DNA sequence and outputs the length of the sequence. The program should make use of <STDIN>. 3  
d) Write a Perl script to count nucleotide frequencies in a DNA sequence. 6
10. Write short notes on any *three* of the following :  $5 \times 3 = 15$   
a) FASTA  
b) Dynamic programming  
c) Gene prediction  
d) EMBOSS.
11. a) Outline the difference in between NW and SW algorithm. Explain the Smith - Waterman algorithm to generate optimal local alignment.  $3 + 6$   
b) What is scoring matrix ? Write a brief note on PAM series.  $2 + 4$