Name : ....

Roll No. : ....

Invigilator's Signature : ....

CS/B.Tech (BT)/SEM-5/BT-503/2010-11
2010-11
BIO-INFORMATICS – 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 :

$10 \times 1 = 10$

i)   The full form for NCBI is

a)   National Consortium for Biotechnology Information

b)   National Consortium for Biology Information

c)   National Centre for Biotechnology Information

d)   National Centre for Biology Information.

ii)   In each line, FASTA sequence contains

a)   100 characters   b)   60 characters

c)   75 characters   d)   80 characters.

iii)   ORF finder gives results for

   a)   naturally occurring reading frame

   b)   all 3 reading frames

   c)   all 2 strands

   d)   all 6 reading frames.

iv)   BLOSUM is a

   a)   Substitution matrix   b)   Alignment matrix

   c)   Both (a) and (b)   d)   None of these.

v)   Which of the following regular expressions would be matched by sequence DWILKDG ?

   a)   D-M-x-[ILV]-x{2}-G

   b)   [DN]-W-x-[ILV]-[RKH]-x-G

   c)   [DN]-W-x{2}-[ILV]-G

   d)   D-W-I[ILMV]-x-K-[GA].

vi)   What is a fingerprint ?

   a)   A protein family discriminator built from a set of regular expressions

   b)   A protein family discriminator built from a set of conserved motifs

   c)   A cluster of protein sequences gathered from a BLAST search

   d)   A cluster of protein sequences gathered from a FASTA search.

vii) Why are colour schemes important in creating and analyzing sequence alignments ?

    a) They look pretty

    b) To make clearer printouts and presentations

    c) To allow you to distinguish conserved residue groups more easily

    d) To allow you to detect active sites of proteins.

viii) Two sequences are said to be homologous if

    a) They have diverged from a common ancestor

    b) Their alignments share 30% identity or more

    c) They belong to the same fold family

    d) They have converged to share similar functional properties.

ix) Bank It is

    a) use of informatics for DNA data bank manipulation

    b) a stand alone multiplatform sequence submission program available in NCBI

    c) a stand alone sequence submission program available in EMBL

    d) a web-based sequence submission program available in NCBI.

x) Which of the following gene finding softwares is available on NCBI ?

    a) Spidey   b) Genscan

    c) ORF finder   d) Genwise.
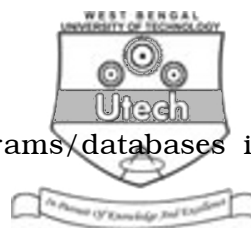
xi) Which of the following search programs/databases is NOT found at the NCBI Website ?

    a) LocusLink   b) PSIPRED

    c) PubMed   d) dbSNP.

xii) Two genes are said to be paralogous

    a) when they are not orthologous

    b) when there are no evidences of gene duplication

    c) when two copies of the duplicated gene and their progeny are found in the evolutionary lineage

    d) none of these.

GROUP – B

(Short Answer Type Questions)

Answer any *three* of the following.   3 × 5 = 15

2. Describe gap penalty. Why has the gap opening penalty a higher value than gap extension penalty ?   2 + 3

3. What is Pfam ? How can you predict a gene and its promoter region ?   1 + 4

4. How can you get repeated sequence in dot plot ? What is the limitation of dot plot ?   3 + 2

5. What is the difference between pair wise alignment and multiple alignment ? What information can we get form these ? State one web-based tool for multiple alignment.   2 + 2 + 1

6. What is a bit score ? Write down the significance of expect value.   2 + 3

GROUP – C

(Long Answer Type Questions)

Answer any *three* of the following.   3 × 15 = 45

7. a) Align the two sequences S1 = ATTAGCTGAC and S2 = TAGCTG, locally by dynamic programming method. Given that scores for match, mismatch and gap are 3, 1 and – 2 respectively.   7

    b) Why is substitution matrix used ? What are the differences between PAM and BLOSUM ? What do you mean by BLOSUM 50 ?   3 + 3 + 2

8. Describe the algorithm of BLAST. What is its difference with Smith-Waterman algorithm ? Why is filtering used ? State the filtering processes used in BLAST search.   7 + 3 + 3 + 2

9. Write the following programs using perl :   5 + 5 + 5

    a) Write a program that will take the DNA sequence, which could be in upper or lower case from a file and print in lower case.

    b) Write a program to reverse transcribe RNA to DNA. RNA sequence remains in the file rna_seq.

    c) Write a program to determine the frequency of nucleotide. The nucleotide sequence remains in file nuc_seq.

10. a) Write down the commands for the following operations in vi editor : 5 × 2

   i) Quit without save.

   ii) Move cursor to end of the file.

   iii) Substitute the word 'weak' by the word 'week' in whole content of file.

   iv) Save the contents from the 5th line through 20th line to another file.

   v) Replace current character with next character typed.

   b) Write down the name of the program used in EMBOSS for following cases : 5 × 1

   i) Visual overview of the distribution of ORFs in the six frames.

   ii) Hydrophobicity profiles of protein.

   iii) Motif finding.

   iv) Display the multiple aligned sequences, with colouring and boxing.

   v) Create profile from a set of multiply aligned sequences.
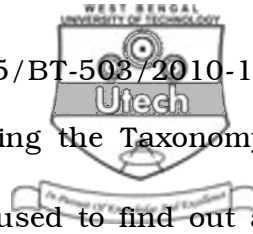
11. How can you get a phylogenetic tree using the Taxonomy browser ? List the limits that should be used to find out a specific article in PubMed. What is the full form of PubMed ? In a Genbank nucleotide sequence, what do you mean by

a) GI no. ?

b) Locus ?

c) Reference ? 6 + 2 + 1 + 6