



Name :

Roll No. :

Invigilator's Signature :

CS/B.Tech(CSE)/SEM-7/CS-704B/2010-11

2010-11

BIOINFORMATICS

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 the following : $10 \times 1 = 10$

i) Below the alignment of four DNA sites for a protein binding is shown :

CAACTG

CAGCTG

CAGGTG

CAGCTT

Which of the following three position-specific score matrices (PSSM) is more likely to be correct ?

a) **Error!)**

b) **Error!)**

c) **Error!)**



ii) Total number of useful amino acids is

- a) 23 b) 20
c) 19 d) none of these.

iii) What is the score for following two sequences (assume
match = + 3, mismatch = - 1, indel = - 3)

C	T	G	G	G	A	T	C	T
C	-	C	G	G	-	T	C	A

- a) 7 b) 6
c) 10 d) none of these.

iv) Hamming or edit distance for following two given words

REASON

SEASON

is

- a) 2 b) 0
c) 5 d) none of these.

v) Which one of the following does not match with the rest
three ?

- a) PDB b) NCBI
c) EMBL d) ORF.



vi) EST means

- a) expressed sequence tags
- b) extreme sequence tags
- c) extended sequence tags
- d) all of these.

vii) UPGMA stands for

- a) unweighted pair group method with arithmetic mean
- b) unpair group method with arithmetic mean
- c) unweighted pair group method with mean arithmetic
- d) none of these.

viii) Gene prediction tool at NCBI is

- a) genescan
- b) gene finder
- c) ORF finder
- d) all of these.

hm?

- # hm?

hm?

hm?

hm?

- # hm?



GROUP – C
(Long Answer Type Questions)

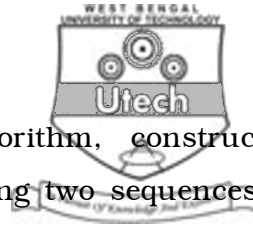
Answer any *three* of the following. $3 \times 15 = 45$

6. a) Outline the role of bioinformatics in drug design and discovery.
- b) Define multiple sequence alignment. What is the goal of multiple sequence alignment ?
- c) Explain simultaneous methods and progressive methods for multiple alignments. $5 + 5 + 5$
7. a) What is dot plot ? Discuss the advantages and disadvantages of dot plot used for. Draw a dot plot for these two sequences :

DILVDEQ

IVQDEQ

- b) Write down steps of the FASTA. $8 + 7$
8. a) Explain Chao-Fasman algorithm with an example.
- b) Why do we create secondary databases ? $10 + 5$
9. a) What are decoding and learning problems in HMM ? How is it used ?
- b) Describe Viterbi algorithm in HMM. $(5 + 5) + 5$



10. a) Using the Needleman-Wunsch algorithm, construct alignment score table for the following two sequences. Assume the scoring parameters + 1 for match score, 0 for mismatch score and – 1 for gap penalty.

SEQUENCE 1 : A T G C G C T A C G T A T T

SEQUENCE 2 : A T G C G C T

- b) What are the differences between Needleman-Wunsch algorithm and Smith-Waterman algorithm with an example. 10 + 5

11. Write short notes on any *two* of the following : $2 \times 7\frac{1}{2}$

- a) Phylogenetic tree
- b) BLOSUM
- c) Protein folding.
