

H 1334

B.E./B.Tech. DEGREE EXAMINATION, MAY/JUNE 2006.

Sixth Semester

Industrial Biotechnology

IB 341 — BIOINFORMATICS

Time : Three hours

Maximum : 100 marks

Answer ALL questions.

PART A --- (10 × 2 = 20 marks)

1. Write the classification of data mining tools in predictive and descriptive category.
2. What are web browsers? Give few examples and their suitability.
3. Expand and explain PSI-BLAST and PHI-BLAST.
4. How can you assess the significance of FASTA searches?
5. Compare PAM and BLOSUM matrices.
6. What is sequence filtering? How can you identify compositional complexity of a sequence?
7. How to calculate the significance of a local alignment using Gumbel's extreme value distribution?
8. Differentiate local and global alignments.
9. Are the beta and alpha chains of the hemoglobin molecule homologue, paralogue and orthologue to each other.
10. How can you find an ORF from a given sequence?

PART B — (5 × 16 = 80 marks)

11. (i) Explain Jukes Cantor and Kimura models of nucleotide substitutions. (4)
- (ii) Explain the various methods of pairwise alignment and write which method gives the optimal alignment. (4)
- (iii) S1 = AATTCGCGTA S2 = TATCGCTACA match+2, mismatch 0, gap -1
Build the complete Dynamic programming table for these strings.
What is the edit distance between S1 and S2?
Give the optimal global alignment between S1 and S2. (8)

12. (a) Explain the following terms : (4 × 4 = 16)
- (i) Affine gap penalty
 - (ii) Log odd score
 - (iii) Smith Waterman Algorithm
 - (iv) Bayesian rule.

Or

- (b) (i) What are the uses of multiple sequence alignment? (2)
- (ii) How do you calculate sum of pairs in an msa? (4)
- (iii) Elaborate on multiple sequence alignment using Sp scores. (10)
13. (a) (i) What are the drawbacks of progressive alignment? (2)
- (ii) How ClustalW and PILEUP performs msa? (4)
- (iii) Describe the steps followed in Genetic Algorithm to do msa. (10)

Or

- (b) (i) How patterns are identified in proteins by profile matrix? (8)
- (ii) Explain Expectation Maximization Algorithm used in msa. (8)
14. (a) What is a BLAST search? How will you do it? Explain the steps used by the BLAST Algorithm. (16)

Or

- (b) (i) What is meant by informative site in a pasimony analysis? How to identify the informative sites? (16)

- (ii) Expand and explain the UPGMA tree and construct a UPGMA tree for the following distances. (10)

	Human	Chimp	Orangutan	Gorilla	Gibbon
Chimp	2				
Orangutan	4	4			
Gorilla	6	6	6		
Gibbon	6	6	6	4	
Monkey	8	8	8	8	8

15. (a) Give brief answers :

(4 × 4 = 16)

- (i) Need for an out group in a phylogenetic tree.
- (ii) Molecular Clock Hypothesis.
- (iii) Rooted Vs Unrooted tree.
- (iv) Mutation rate in Nuclear and mitochondrial sequences.

Or

- (b) Explain briefly the following :

- (i) Gene prediction by HMM and Neural Network methods. (8)
- (ii) Shotgun strategy for complete genome sequencing and bioinformatics tools to assemble the data. (8)