STELLA MARIS COLLEGE (AUTONOMOUS) CHENNAI 600 086
(For candidates admitted during the academic year 2004– 05 & thereafter)
SUBJECT CODE: BT/MO/IB64

B.Sc. DEGREE EXAMINATION, APRIL 2008
BRANCH V (a) – PLANT BIOLOGY AND PLANT BIOTECHNOLOGY
SIXTH SEMESTER

COURSE : MAJOR – OPTIONAL
PAPER : INTRODUCTION TO BIOINFORMATICS
TIME : 3 HOURS MAX. MARKS: 100

SECTION – A

I FILL IN THE BLANKS: (6 marks)

1. Maps produced by the cutting of rare restriction enzymes are called …….
2. Hash table look up is used in the …….. algorithm.
3. Chou-Fasman table is used in the prediction of …….
4. BLOSUM matrices have been constructed from …….. database.
5. An example of a protein structure database is ……..
6. The full form of KEGG is ……

II. **MULTIPLE CHOICES**: **(6 marks)**

1. Which of these regular expressions will be matched by the subsequence "EWILYHG"
   a) E [WYF] I [LV]XH[AG] b) E{WYF}[IL]LXHG
   c) Both a and b d) EW[IL] YHG
2. Which of these give an initial penalty for gap initiation and a lower penalty for gap extension
   a) Constant gap penalty b) Affined gap penalty
   c) Initiation penalty d) Extension penalty
3. The banding patterns of a chromosome is seen in a
   a) Cytogenetic map b) Genetic map
   c) Radiation map d) Recombination map
4. These served as landmarks in constructing the physical map of the human genome
   a) Sequence Tagged Sites b) Contigs
   c) Single nucleotide polymorphisms d) Exon intron boundaries
5. Example(s) of super secondary structures in proteins
   a) Beta barrel b) Greek key c) a and b d) None of the above
6. GSS division of GenBank refers to
   a) Genome Survey Sequences b) Genome Sequence Summary
   c) Gene Split Sequence d) Gene Scoring Scheme

III. **STATE WHETHER TRUE OR FALSE: (6 marks)**

1. BLOSUM 80 is used highly divergent sequences.
2. Profiles are quantitative as compared to sequence motifs.
3. Lead compound generally has all the desired properties.
4. INDELS will always lead to frame shift mutations.
5. Dynamic programming can be used to find edit distances of two sequences.
6. A heuristic algorithm does not give an exact solution.

..2..

IV. **ANSWER IN 50 WORDS EACH:** **(18 marks)**

1. What are NJ trees?
2. What are pre clinical and clinical trials in the drug discovery pipeline.
3. What are centromeres?
4. What is a Ramachandran plot?
5. What are DNA chips?
6. What are profiles?

SECTION – B

ANSWER ANY FOUR NOT EXCEEDING 300 WORDS **(4x6=24)**

1. What is partial sequencing? How is it used in the mapping of a genome?
2. Explain Smith Waterman algorithm.
3. Align two sequences RTYA and PTALA using a global alignment approach. Give all possible alignments.
4. Discuss the construction of PAM substitution matrix. Why is it also called a log odds matrix?
5. How do you normalize scores in BLAST. If you are scoring sequences, which are very related which BLOSUM matrix would you use? Justify your answer. A BLAST hit has an e-value of 0.0. What is its significance?
6. Write a short note on any one organism specific database.
7. What is DNA computing?

SECTION – C

ANSWER ANY TWO NOT EXCEEDING 1200 WORDS **(2x20=40)**

1. What is cloning? What are the different cloning vectors used in genome sequencing projects. How is a cDNA library and a cDNA map constructed?
2. Write in detail any one method of Tree construction.
3. What are the tools/resources used for comparison of two PDB structures. Comment on the principle behind the structure comparison. What are protein structure classification databases?
4. Write short notes on
   a) identification and optimization of lead compounds in a drug discovery pipeline.
   b. Features of tRNA and its secondary structure prediction.

****