Assignment – III

(Sequence Alignment)

Deadline: 10th April

1. You are given 2 nucleotide sequences:
   GGCTGCAACTAGCTC
   GGGTAAGCTTGC
   and the transition-transversion scoring matrix (expressed in similarity):

	A	C	G	T
A	4	-1	1	-1
C	-1	4	-1	1
G	1	-1	4	-1
T	-1	1	-1	4

   and gap penalty -3.

   Carry out the global and local alignment (dynamic programming algorithm), and indicate the final similarity score and the best alignment.

2. Identify the dinucleotide CA repeat region and the score in the following sequence:
   TGGCACACTCACACCACACAGACAGTTA

3. When would you encounter a situation for using DP for overlap regions? How are the boundary conditions and recursive relations different from that for global alignment?

4. What is the advantage of using affine gap scores?

5. Give the time and space complexity of DP. Under what conditions is time an issue and under what conditions would space be a problem?

6. Describe the construction of Nucleic acid PAM scoring matrices.

7. Take any gene sequence and its corresponding protein sequence and perform databases searches with both these sequences. Which of these two searches identifies more significant matches? Give reasons.

8. What is the difference in the working of PSI-BLAST and BLAST programs?

9. (i) In BLAST database search algorithm, the match/mismatch ratio for comparing nucleotide sequences is chosen to be large for highly conserved sequences, while it is small for divergent sequences. Give reasons, why?
   (ii) Give the BLAST nucleotide substitution matrix for comparing sequences that are 95% conserved.

10. In BLOSUM62 matrix, a conserved Tryptophan position has score S(W,W) = 11, but a conserved Leucine position has score S(L,L) = 4. Give at least one reason why these values differ.

11. Construct the scoring scheme for identifying DNA sequences that exhibit at least 65% identity. Assume background frequency 0.25, for each of nucleotides and assume equiprobability for mismatch.