**Assignment – IV**

**(Sequence Alignment)**

**Deadline: 14-09-20**

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.

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

TGGCACACTCACACCACACAGACAGTTA

1. 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?

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

1. Does a Global Alignment Algorithm always give a global alignment and a Local Alignment Algorithm always give a local alignment? Give reasons for your answer.
2. Give the time and space complexity of DP. Under what conditions is time an issue and under what conditions would space be a problem?
3. Find out the size of protein database, UniProt, and the nucleotide database, GenBank, and compute no. of matrix cells needed to be computed by using DP and the time taken for database search against Uniprot & GenBank. Assume query sequence of length 1000 bases and computation time as 10 million matrix cells per second.