BE7-R3: APPLIED BIOINFORMATICS

NOTE:

1. **Answer question 1 and any FOUR questions from 2 to 7.**
2. **Parts of the same question should be answered together and in the same sequence.**

Time: 3 Hours **Total Marks: 100**

1.

a) Define the terms "Transcription" and "Translation" in the context of protein synthesis.

b) Consider the DNA string 5' TTCCAACGGCCATGA 3'. Write all the 5' substrings ending with A.

c) What is the order of computational complexity in pairwise sequence comparisons?

d) Why is dynamic programming technique not considered for large scale database homology search such as "BLAST"?

e) How is the degree of dependence defined in a simple markov chain model?

f) What are the disadvantages of using higher order markov model?

g) Outline the purpose or relevance of performing "Multiple Sequence Alignment".

(7x4)

2.

a) How are bioinformatics tools and techniques helpful in analyzing raw sequence and structure data?

b) Compare and contrast the genomic structure of Prokaryote and Eukaryote.

c) Define low complexity sequence and why should it be masked?

(8+6+4)

3.

a) Provide the basic dynamic programming algorithm for the local and global pair-wise sequence alignment problem.

b) Consider the problem of aligning two DNA sequence of E.coli. These sequences are GGGTGATTAGCT and GCTGATATAGCT. Derive a Global similarity alignment for the above two sequence using the scoring scheme S(a,a) = +1; s(a,b) = -1 if $a \neq b$; and gap = 2.

c) Write the consensus for the following multiple alignment sequences:

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

(6+8+4)

4.

a) List at least the four versions of the popular "Blast" homology search tool.

b) Explain the basic algorithmic steps of "BLAST".

c) How is the statistical significance of the "Blast hit" computed". What are the parameters that affect such computations?

(4+6+8)

5.

a) Let the state symbols for the positive model be given as A+, T+, G+ and C+ and for the negative model be given as A-, T-, G- and C-. In a typical Markov Model framework, transitions can take place in two ways i.e. within and between the models. Draw the state transition diagram separately for these transitions.

b) Explain how a simple prediction strategy can be developed using a first order Markov Chain model for discriminating a biologically important functional site.

(8+10)

6.

a) Explain the hidden markov model framework. How does it differ from a markov chain?

b) Explain the viterbi algorithm used in the popular “HMM model”.

c) How is Hidden markov model framework applied to multiple sequence alignment or gene prediction problem?

(4+6+8)

7.

a) How is a “profile” distinct from a “motif”?

b) Discuss progressive alignment method employed in Multiple sequence alignment problem.

c) Explain, “sum of pairs” scoring method in the context of multiple sequence alignment.

(4+8+6)