



**Answer the following questions (Total Marks: 60 M):**

1. [5 M] What is bioinformatics? What is its ultimate goal? List the three goals of the bioinformatics field?
2. [10 M] There are two sequences: A = TACTAA, B = TAATA. Assume that Match score equals +1, Mismatch score equals -1, and Indel score equals -2. Calculate the best similarity sequences for the following methods:
  - a. Local alignment method.
  - b. Global alignment method.
3. [10 M] For the following BLAST hit. Given the following parameters:  
 Query length: 150,  
 $\lambda = 1.37$ ,  $K = 0.711$   
 Average Sequence length in database: 270  
 Number of sequences in database: 4,554,026
 

ACGTCGATCGAGCT  
 ||||| |||||  
 AGGTCGTC-GAGGT

  - a. Calculate the S, S' and E.
  - b. What will be the minimal score in order to achieve a significant E value ( $e^{-6} \sim 10^{-6}$ )?
4. [5 M] What is the general idea of the BLAST algorithm? Write the main steps of this algorithm?
5. [10 M] Compare between PAM and BLOSUM? Write the main steps of the searching process for remote homology using PSI-BLAST?
6. [10 M] There are four sequences as shown below:
 

Sequence a: ACGCGTTGGGCGATGGCAAC

Sequence b: ACGCGTTGGGCGACGGTAAT

Sequence c: ACGCATTGAATGATGATAAT

Sequence d: ACACATTGAGTGTGATAATA

Using the UPGMA distance method to build a dendrogram for these four sequences.

7. [10 M] Finding TF targets using a bioinformatics approach that has binding motif is known. Starting from the following set of aligned motifs and length  $l=7$ :
 

Seq 1	AAAGCCC	Seq 4	CTATCCC	Seq 7	CTATCCC
Seq 2	CTATCCA	Seq 5	GTATCCC	Seq 8	CTATCCC
Seq 3	CTATCCC	Seq 6	CTATCCC	Seq 9	TTATCTG

  - a. Determine the Position Weight Matrix (PWM).
  - b. Calculate  $\text{Pr}(\text{CTAATCCG})$ .

*With my best wishes*  
 Assoc. Prof. Dr. Mohammed Elmagdy