Ain Shams University
Faculty of Science
Department of Mathematics
Computer Science Division
Total Marks: **105 Marks**
Date: **August 20, 2023**

كلية العلوم

Level: **Four** - Semester: **Summer 2023**
Course Code: **COMP 402**
Course Name: **Bioinformatics**
Programs: **CS (Special & Double)**
Time Allowed: **3 Hours**
Number of pages: **2**

Answer the following questions

Question 1 [15 pts]:

1. [2 pts] Mention three differences between DNA and RNA.
2. [4 pts] **Describe shortly**, the follow of genetic information due to the "*Central Dogma of Molecular Biology*".
3. [8 pts] **Put (√) or (×) for each statement. <u>Give reasons or examples for the correct statements or Correct the false ones</u>**
   a. The organism's complete set of DNA is called the gene.
   b. The DNA sequence consists of 20 amino acids.
   c. The distance between a real motif and the consensus sequence is generally less than the distance between two real motifs.
   d. The Median String problem is not equivalent to the Motif finding problem.

Question 2 [20 pts]:

1. [4 pts] What is Bioinformatics? And what are its goals?
2. [3 pts] What is meant by the profile matrix, motif consensus and scoring function of the motif?
3. [6 pts] What is the motif-finding (15,4) problem? What are their complications? Write the mathematical formulation of the implanted motif problem.
4. [7 pts] Write the **Brute Force Algorithm** for solving the motif-finding (15,4) problem. Discuss its running time. Is it practical to use it with real biological data?

Question 3 [25 pts]:

1. [4 pts] Illustrate the workflow and the bottlenecks of Bioinformatics.
2. [5 pts] Define the Hamming distance and the edit distance. Given the two sequences: *CACTG* and *GCACT*, compute Hamming distance and the edit distance between both sequences.
3. [8 pts] Write the mathematical formulation for the "*Change Problem*" of converting an amount of money into given denominations. Write a **recursive algorithm** for solving this problem and a **dynamic-programming algorithm** for solving this problem.
4. [8 pts] Compare between the above two algorithms. Use this example to illustrate the comparison: ($M = 8$ and $C = (1,2,5)$).

P.T.O.

Question 4 [20 pts]: Match 1

Given two sequences v= TACGGGTAT and w= GGACGTACG and assume that the score of the match is 1, mismatch is -2, insertion is -1, and deletion is -2. Assume v is the source sequence which constitutes the columns of the matrix.

1. [3 pts] Write the recurrence relation to compute the optimal global alignment.
2. [3 pts] Fill out the global alignment matrix between v and w.
3. [1 pts] What is the score of the optimal global alignment between v and w.
4. [1 pts] How many optimal global alignments between v and w.
5. [1 pts] List all optimal global alignments between v and w.
6. [11 pts] Add a constraint on the recurrence relation of the global alignment so that the sequence v is **entirely** aligned to a **subsequence** of w, i.e., each element in v is aligned to either indel or an element of w, but not necessarily each element in w is aligned to something.
   a. [5 pts] Write the modified recurrence relation (and the scoring system if needed).
   b. [3 pts] Apply the modified algorithm and write the alignment matrix with backtracking to align v and w.
   c. [3 pts] List all constrained global alignments between v and w.

(handwritten)
v T A C G G G T A T
w G G A C G T A C G

Question 5 [25 pts]:

1. [15 pts] Assume there is a permutation of length 5 such that the distance to the identity permutation using the breakpoint graph is less that the distance using the 2-break reversal sorting algorithm.
   a. [7 pts] Write a permutation of length 5 so that the above observation is satisfied.
   b. [5 pts] Draw all breakpoint graphs to transform the permutation into the trivial graph. Report the number of cycles for each graph.
   c. [3 pts] Write down all intermediate permutations in the reversal soring algorithm. Report the number of breakpoints for all intermediate permutations.
2. [5 pts] Assume that the permutation p = -1 +3 -6 -2 +7 -4 +5.
   1. [1 pts] What is the lower bound distance between p and the identity permutation?
   2. [4 pts] Apply the greedy algorithm to transform p into the identity permutation. Report all intermediate permutations and report the total number of reversal operations that were performed.
3. [5 pts] what could be a modification to the reversal sorting algorithm to transform the sentence "the sky is blue" to "blue is the sky"? apply the algorithm to transform the first sentence to the second one.

(handwritten) blue VS the the

With our best wishes
Dr. Mohammad Fakhry & Dr. Mohammad Hashim

(handwritten bottom)
W T - A T G G G C A T -
W G C A T G - - C A G G
V G - T A T G G G C A - T