

In the Name of the Most Knowledgeable

Faculty of Computer Engineering

Sharif University of Technology

Introduction to Bioinformatics

Midterm Exam - Fall 2022

Duration: 2.5 hours

Date: 10/08/2022

Question 1: True or False

Determine whether the following statements are true or false, and provide reasons for your answers:

- a) In the transcription process, **RNA polymerase** binds to the start site of a gene with the help of transcription factors.
 - b) In the case of a gene containing three exons, deleting exon 2 is generally more dangerous than deleting exon 3.
 - c) Proteins are functional units in cells, while all RNAs are non-coding and do not contribute to cellular functions.
-

Question 2: Short Answer Questions

Answer the following:

- a) Identify the labeled sections of the diagram and explain their functions.
 - b) What is the reason for the formation of **Okazaki fragments** during cell division?
 - c) Suppose two genes have self-activation and mutual inhibition relationships. Write the corresponding differential equations for these genes and analyze the gene expression behavior under these conditions.
-

Question 3: RNA Folding

Using an algorithm for RNA folding taught in the course, determine the most stable parentheses (valid parenthesis notation) for the following RNA sequence:

Input: UCAAGAAUC

Note: Writing the algorithm's steps is necessary. Only providing the final answer will not receive full credit.

Question 4: Global Sequence Alignment

Consider two sequences **S = TCCTT** and **T = GGAA**. The alignment table has been completed using dynamic programming, as shown below:

	1	2	3	4	5
0	0	0	0	0	0
4	4	4	0	0	2
4	2	5	5	0	3
4	7	10	5	0	4
11	4	7	2	0	5
18	11	4	0	0	6

Answer the following:

- How were the gap penalties calculated in this algorithm? Why?
 - Is this a global or local alignment problem?
 - Write the similarity matrix for the sequences, as in the example where the conversion score for A to C is 7.
-

Question 5: Longest Repeating Subsequence

Design an algorithm using dynamic programming to find the longest repeated non-overlapping subsequence in a string **S** in $O(n^2)$ time.

Example:

Input: S = AACAAACAACA

Output: AACAA

Question 6: Distance Matrix and Phylogenetic Trees

The following distance matrix is provided for four species:

	A	B	C	D
A	0	3	2	5
B	3	0	3	5
C	2	3	0	5
D	5	5	5	0

Answer the following:

- a) Is this distance matrix ultrametric?
 - b) Construct a phylogenetic tree using the **UPGMA algorithm**.
-

Question 7: Additive Phylogeny

The following distance matrix is provided for four species:

	A	B	C	D
A	0	4	10	9
B	4	0	8	7
C	10	8	0	9
D	9	7	9	0

If the matrix allows for constructing a tree using the **additive phylogeny algorithm**, construct the tree and write all the steps in detail. If not, use another algorithm you have learned to reconstruct the tree.