North East University Bangladesh

Department of Computer Science and Engineering
Mid Semester Examination, Summer-2022
Program: BSc.(Engg.) in CSE
Course Code: CSE 449
Course Title: Bioinformatics

Marks: 30 Time: 90 minutes

[NB: FIGURES IN THE MARGIN INDICATE FULL MARKS]

1. Answer any three questions 3x2=6

a) What do you understand by the biological information of a human body?
b) What is the responsibility of the messenger Ribonucleic Acid (mRNA) in the cell?
c) Write at least two differences between eukaryotic and prokaryotic cells.
d) What does the genetic code mean?
e) Write True/False. If False, write the correct answer.
   i. The nucleic acid polymer is called a nucleotide.
   ii. A group of retroviruses can perform RNA replication.

2. Answer any three questions

James Watson and Francis Crick originally published the central dogma theory in 1957. At the time, scientists tried identifying the connection between Deoxyribonucleic Acid (DNA), Ribonucleic Acid (RNA), and protein. Scientists thought such cases that perhaps –

i. DNA was converted to RNA, and
ii. There was still a question of how two-dimensional DNA created three-dimensional proteins.

a) Why is central dogma theory much significant? Explain, with an example, the case study (i). 1.5+1.5=3
b) Elaborate and analyze the case (ii). 3
c) What is nucleic acid? Why do multiple codons represent an amino acid? 1+2=3
d) *The genetic code is degenerative* – explain the fact of the central dogma. 3
e) Identify the codons from the following RNA sequence. Write the name of at least three amino acids from that corresponding codon. 1+2=3

5′ *AGCCGGAGUUUUUUGACUCAUUGAAAUC* 3′

3. Answer any three questions

Very short or similar sequences can be aligned by hand. However, most interesting problems require the alignment of lengthy, highly variable, or excessively numerous sequences that cannot be aligned solely by human effort. Instead, human knowledge is applied in constructing algorithms to produce high-quality sequence alignments and occasionally in adjusting the final results to reflect patterns that are difficult to represent algorithmically. Computational approaches to sequence alignment generally fall into two categories: *global* and *local alignments*. Furthermore, If two sequences in an alignment share a common ancestor, mismatches can be interpreted as point mutations.

a) i. What does sequence alignment mean? Why is it important? 1+2=3
   ii. How do you distinguish between FASTA and BLAST? 1

iii. Find the reverse complement of the following sequence – 1

$5'\ TGCATTAGC\ 3'$

b) When is local alignment effective in aligning the sequence? Find and identify the alignment score by applying the Smith-Waterman method for the following sequences – 2+3=5

Lab Sequence (LS)	$5'\ AAGTCA\ 3'$
Reference Sequence (RS)	$5'\ AAATC\ 3'$

Consider match $= +1$, mismatch $= -1$, and gap penalty is -2.

c) What is the concept of global alignment? Complete the matrix and find the global alignment score between two sequences by applying the appropriate algorithm's initialization, table filling, and traceback process. Consider match $= +1$, mismatch $= -1$, and gap penalty is -2. 1+4=5

	0	A	G	C	G	T	A	G
0	0	−2		−6		−10	−12	
C		−1		−3			−9	
T						−6		
C	−6			1				
G		−7						−4
T								
C	−12		−8					

d) i. Count the point mutation between $LS = TGCATAT$, and $RS = ATCCGAT$. 2

ii. Calculate how many minimum operations are required to convert RS to LS. Note that the operation includes *insertion*, *deletion*, and *substitution*. 3

e) Consider the reference and lab sequences $ATATTACG$ and $ATCG$, respectively. Assume that the premium match is $+1$ and that mismatch and gap penalties are 0. Count how many gaps would be occurred between the sequence alignments. 5