MASTER OF COMPUTER SC. & ENGG. 2ND SEM. EXAMINATION – 2024

BIOINFORMATICS

Time: Three Hours **Full Marks: 100**

Answer any FOUR questions

1. a) Briefly Describe the BLAST algorithm. Explain with respect to the following example (consider default word length as 4):
Input sequence: AILVPTV
Database sequence: MVQGWALYDFLKCRAILVGTVIAML

b) Briefly discuss n-star consensus strategy in AMS algorithms for prediction of PTM sites.

c) What is molecular docking? How different surface representation methods identify the regions of interest. Describe any two surface representation methods in brief. What is the basic working principle of the DOCK algorithm?

10+5+10 = 25

2. a) Use UPGMA to construct a phylogenetic tree using the following distance matrix:

	A	B	C	D	E	F	G
A	0.0						
B	19.00	0.0					
C	27.00	31.00	0.0				
D	8.00	18.00	26.00	0.0			
E	33.00	36.00	41.00	31.00	0.0		
F	18.00	1.00	32.00	17.00	35.00	0.0	
G	13.00	13.00	29.00	14.00	28.00	12.00	0.0

b) Consider the following DNA sequences. Construct the corresponding Position Frequency Matrix (PFM) and the resulting Position Probability Matrix (PPM).

GAGGTAAAC
TCCGTAAGT
CAGGTTGGA
ACAGTCAGT
TAGGTCATT
TAGGTACTG
ATGGTAACT
CAGGTATAC
TGTGTGAGT
AAGGTAAGT

15+10=25

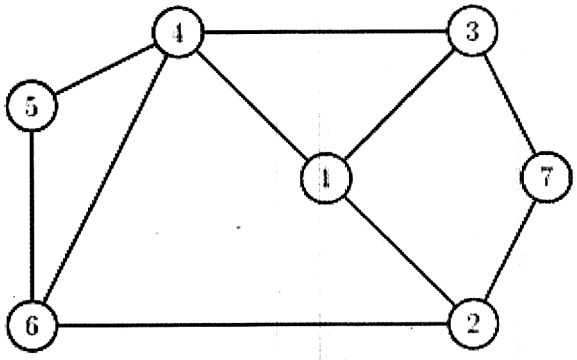
Using the Needleman and Wunsch dynamic programming method, find the optimal global alignment(s) for the following sequences. Use the BLOSUM-62 scoring matrix given below and linear gap penalty as 4.
Sequence 1 = THISLINE
Sequence 2 = ISALIGNED

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W
C	9																			
S	-1	4																		
T	-1	1	5																	
P	-3	-1	-1	7																
A	0	1	0	-1	4															
G	-3	0	-2	-2	0	6														
N	-3	1	0	-2	-2	0	6													
D	-3	0	-1	-1	-2	-1	1	6												
E	-4	0	-1	-1	-1	-2	0	2	5											
Q	-3	0	-1	-1	-1	-2	0	0	2	5										
H	-3	-1	-2	-2	-2	-2	1	-1	0	0	8									
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5								
K	-3	0	-1	-1	-1	-2	0	-1	1	1	-1	2	5							
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5						
I	-1	-2	-1	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4					
L	-1	-2	-1	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	4				
V	-1	-2	0	-2	0	-3	-3	-3	-2	-2	-3	-3	-2	1	3	1	4			
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6		
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	7	
W	-2	-3	-2	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11

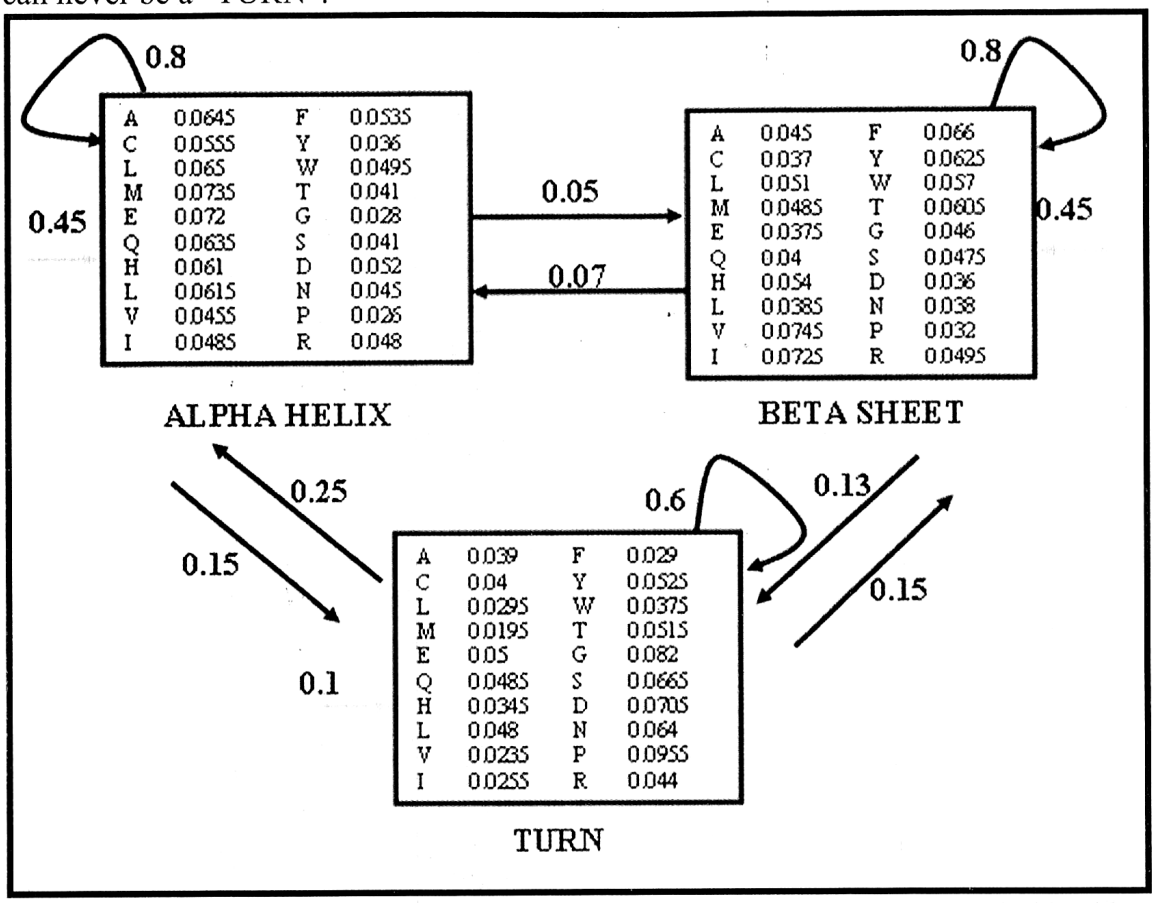
25

a) What is Protein Secondary Structure Prediction (PSSP)? Briefly discuss the PSIPRED method for PSSP.

b) Consider the following graph and estimate the *degree distribution* and *clustering coefficient* for all the nodes in the graph.



c) Given the following probability distribution, use Hidden Markov Model to estimate the optimum secondary structure annotation for a primary sequence: **TAS.** T is the starting Amino Acid in the sequence, and assume that the first two Amino Acids in the sequence can never be a "TURN".



5+10 +10 = 25

Write short notes on the following (any 5):

a) DOT Plot
b) Central Dogma of Molecular Biology
c) BLOSUM
d) Scale Free Network
e) Ubiquitination
f) KEGG

5x5 = 25