U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

September / October 2023 Supplementary Examinations

Programme: B.E.	**Semester: VI**
Branch: Biotechnology	**Duration: 3 hrs.**
Course Code: 19BT6DCBIN	**Max Marks: 100**
Course: Bioinformatics	**Date: 19.09.2023**

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

UNIT - I

1 a) Explicate the features of protein structure databases **06**

b) With a suitable illustration, deliberate the features of Genbank flat file format **08**

c) Discuss the features of OMIM database. **06**

UNIT - II

2 a) Construct an alignment with Needlman & wunch algorithm using dynamic programing. Evaluate your score to determine the optimal alignment for the following sequences. **10**

Sequence1= **CAAAGTCTACGA**, Sequence2= **AAACGAGCTACG**

Sl. no.	Parameter	score
1	Identity	+3
2	Mismatch	-1
3	Gap creation	-2

b) Dot-plots provide a visual representation of sequence similarity. Substantiate the statement for the following example and interpret your result. **05**

Top sequence: GAGACCCTGATTACGAT
Left sequence: GAGA TGATTACGAT

c) Elaborate the steps involved in FASTA algorithm and comment on its statistical significance. **05**

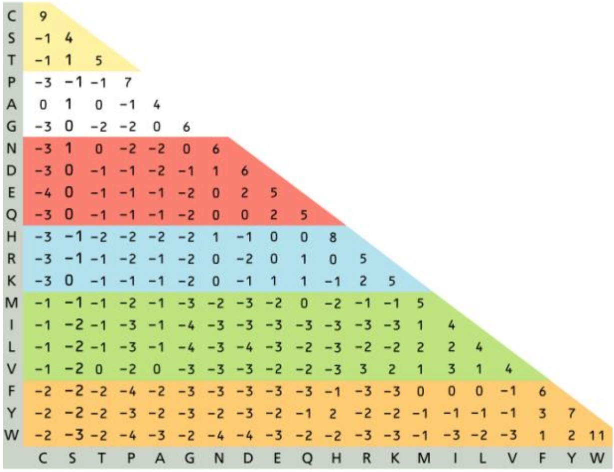
OR

3 a) Illustrate with an example, the steps involved in MSA using progressive alignment method **06**

b) Determine the score for the following sequence alignment considering the substitution matrix and gap penalty scores. **04**

```
V - E T D M W T P L T R L A S L R Q H
L A Q T T L - - K A H T L T S L G T H
```

Sl no.	Parameter	score
1	Gap creation	-4
2	Gap extension	-1



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
	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W

c) Construct an alignment using Smith & Waterman algorithm. Evaluate your score to determine the optimal alignment for the following sequences. **10**

Sequence 1= **ACCTGGACACGCT** , Sequence 2 = **ACACTCCGCATCG**

Sl no.	Parameter	score
1	Identity	+4
2	Mismatch	-1
3	Gap creation	-2

UNIT - III

4 a) Apply UPGMA method to make a phylogenetic tree for the following distances between six taxa and determine the branch lengths **10**

	A	B	C	D	E	F
A	-					
B	2	-				
C	4	4	-			
D	6	6	6	-		
E	6	6	6	4	-	
F	8	8	8	8	8	-

b) Construction a Position specific scoring matrix for the following multiple alignment of nucleotide sequences. Determine the probability of a new sequence ACGAAG fit into the matrix? **10**

Position	1	2	3	4	5	6
Sequence 1	T	C	A	T	A	C
Sequence 2	A	G	G	C	G	G
Sequence 3	A	C	T	A	A	G
Sequence 4	C	T	G	A	A	C
Sequence 5	T	C	G	A	T	G

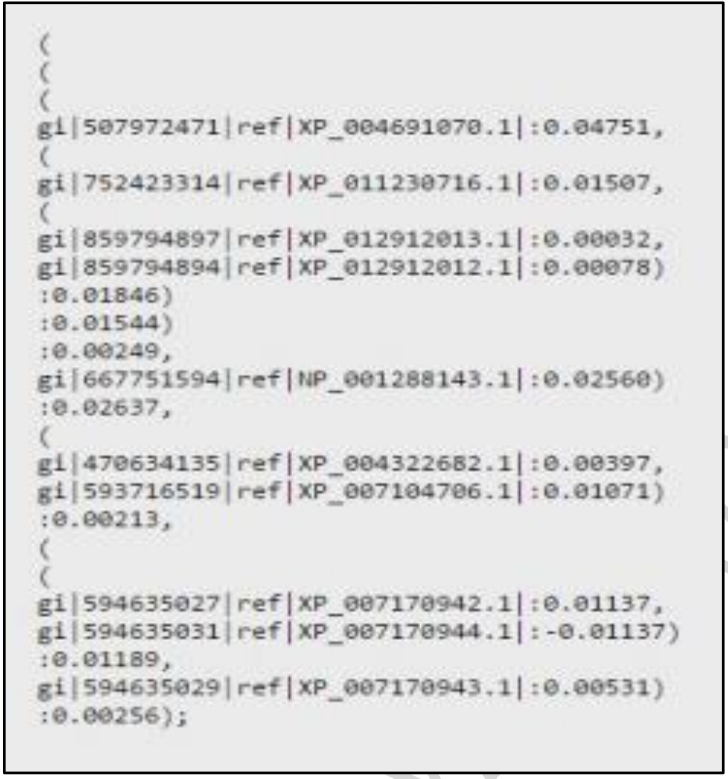
OR

5 a) Construct a Hidden Markov model to represent a multiple sequence alignment. **05**

b) For the following distances between five taxa, construct a phylogenetic tree using Neighbourhood joining method and determine the branch lengths. **10**

	A	B	C	D	E
A	-	12	14	14	15
B	-	-	12	12	13
C	-	-	-	6	7
D	-	-	-	-	3
E	-	-	-	-	-

c) Generate a cladogram tree for the following guide tree data output. **05**



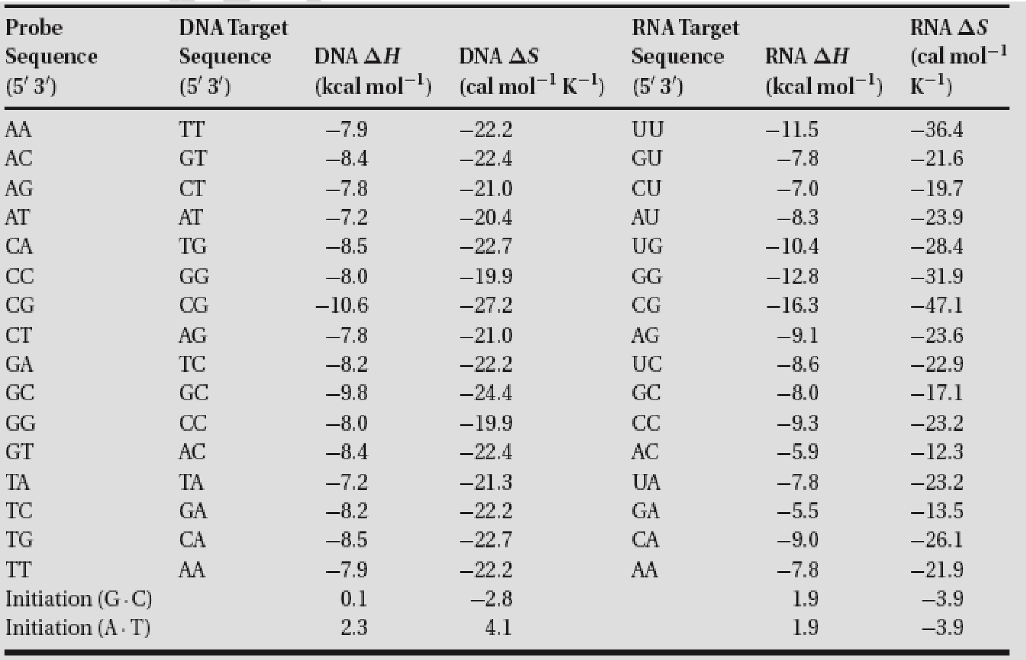
```
(
(
(
gi|507972471|ref|XP_004691070.1|:0.04751,
(
gi|752423314|ref|XP_011230716.1|:0.01507,
(
gi|859794897|ref|XP_012912013.1|:0.00032,
gi|859794894|ref|XP_012912012.1|:0.00078)
:0.01846)
:0.01544)
:0.00249,
gi|667751594|ref|NP_001288143.1|:0.02560)
:0.02637,
(
gi|470634135|ref|XP_004322682.1|:0.00397,
gi|593716519|ref|XP_007104706.1|:0.01071)
:0.00213,
(
(
gi|594635027|ref|XP_007170942.1|:0.01137,
gi|594635031|ref|XP_007170944.1|:-0.01137)
:0.01189,
gi|594635029|ref|XP_007170943.1|:0.00531)
:0.00256);
```

UNIT - IV

6 a) Calculate the thermodynamics of the following oligo nucleotide probe and predict its melting temperature with 1pM target concentration. Analyze your results and conclude if the probe can be selected or not. If *T*m within ± 2^0 of 60^0C. (Refer table) **10**

5' CTAAACCCAGTATATACTAACTGGTTTTTT 3'

Table1: Thermodynamics properties



Probe Sequence (5′ 3′)	DNA Target Sequence (5′ 3′)	DNA ΔH (kcal mol^{-1})	DNA ΔS (cal mol^{-1} K^{-1})	RNA Target Sequence (5′ 3′)	RNA ΔH (kcal mol^{-1})	RNA ΔS (cal mol^{-1} K^{-1})
AA	TT	−7.9	−22.2	UU	−11.5	−36.4
AC	GT	−8.4	−22.4	GU	−7.8	−21.6
AG	CT	−7.8	−21.0	CU	−7.0	−19.7
AT	AT	−7.2	−20.4	AU	−8.3	−23.9
CA	TG	−8.5	−22.7	UG	−10.4	−28.4
CC	GG	−8.0	−19.9	GG	−12.8	−31.9
CG	CG	−10.6	−27.2	CG	−16.3	−47.1
CT	AG	−7.8	−21.0	AG	−9.1	−23.6
GA	TC	−8.2	−22.2	UC	−8.6	−22.9
GC	GC	−9.8	−24.4	GC	−8.0	−17.1
GG	CC	−8.0	−19.9	CC	−9.3	−23.2
GT	AC	−8.4	−22.4	AC	−5.9	−12.3
TA	TA	−7.2	−21.3	UA	−7.8	−23.2
TC	GA	−8.2	−22.2	GA	−5.5	−13.5
TG	CA	−8.5	−22.7	CA	−9.0	−26.1
TT	AA	−7.9	−22.2	AA	−7.8	−21.9
Initiation (G · C)		0.1	−2.8		1.9	−3.9
Initiation (A · T)		2.3	4.1		1.9	−3.9

b) A 11 kb circular plasmid pRIT455 is digested with three restriction enzymes *ECORI*, *BamHI and HindIII* individually and in combination. The resulting fragment sizes are determined by means of electrophoresis. The results are as follows: **10**

Sl No	Restriction enzymes	Fragment sizes (kb)
1	*ECORI*	8.5, 2.0, 0.5
2	*HindIII*	5.6, 5.4
3	*BamHI*	6.5, 4.5
4	*BamHI and ECORI*	4.5, 2.2, 2.0, 1.8, 0.5
5	*ECORI and HindIII*	4.3, 4.2, 1.2, 0.8, 0.5
6	*HindIII and BamHI*	3.4, 3.1, 2.5, 2.0

Draw a restriction map based on these results.

UNIT - V

7 a) Explicate the Quantitative structure activity relationship with reference to drug discovery **05**

b) Discuss the molecular dynamics simulation model with a neat flowchart. **05**

c) Deliberate the scoring functions involved in grid calculation and binding energy minimization. **05**

d) Demonstrate the identification of new drug candidates through pharmacophore prediction. **05**

******