INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

Mid-Spring Semester 2022-23

Date of Examination: 20-02-2023 Session: AN Duration: 2 hrs Full Marks: 60
Subject No: CS61060 Subject: Computational Biophysics: Algorithms to Applications
Department/Center/School: Department of Computer Science and Engineering
Specific charts, graph paper, log book etc., required: None
Special Instructions (if any): *(1) Answer all the questions. (2) In case of reasonable doubt, make practical assumptions and write that on your answer script. (3) The parts of each question must answered be together.*

---

1. (a) Write down an algorithm for the global pairwise sequence alignment. What is the time and space complexity of your algorithm?

(b) Apply your algorithm to align following two sequences. Explicitly show the final scoring and trace-back matrix for the alignment process. What is your alignment score? Match/Mismatch score is as per BLOSUM62 matrix (provided at the end), gap opening penalty (including one gap extension) is -5 and gap extension penalty is -1.

Input in FASTA format

```
>Protein Sequence1
EAVTQESALTSPGTVTLC

>Protein Sequence2
EVTLVESDSVKPGSLKLC
```

Marks: (8+2)+(5+4+1)=20

---

2. (a) Write down an algorithm/pseudo code that takes (read) a multiple sequence alignment as input and computes the Henikoff weight. Apply your algorithm for the following multiple sequence alignment to compute the Henikoff weight of each of the input sequence. Assume X is an unknown amino acid.

```
1DAP_1|      IAKXQPDMDLVGIX
12AS_1|      IAKXQRQISFVKSH
1EZG_1|      KATACTNSSGCPGH
1IND_1|      GGTXNNRAPGVPAR
1IND_2|      GGGFTFYSASVKGR
```

(b) Analyse the time complexity of your algorithm.

Marks: (8+10)+2=20

---

3. (a) Describe the features you wish to compute for predicting secondary structure from protein sequence.

(b) What is the Q3 error in protein secondary structure prediction?

(c) Draw the geometry required for the formation of the protein main chain hydrogen bonding in presence and absence of hydrogen atom.

Marks: 3+2+5=10

4. (a) Define protein-protein docking problem.

(b) If docking of two protein molecules is called as a binary docking then what is the minimum number of such binary docking required for generating a homo octamer protein complex from a single protein monomer (subunit). Explain. (**Note:** A homomer is a protein complex where all the protein subunits are same.)

(c) What are the advantages of the geometric hashing based protein docking decoy generation?

Marks: 3+3+4=10

BLOSUM62 Matrix

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