

BIO213-Introduction to Quantitative Biology
Assignment 1
Kesar Shrivastava, 2019051

Question 1)

Given sequences: 1. ATCAGAGTA

2. TTCAGTA

For all the generated matrices given below note that the first sequence is at the top and the second is at the left of the matrix.

a) The generated dynamic programming matrix is:

0	-1	-2	-3	-4	-5	-6	-7	-8	-9
-1	-1	1	0	-1	-2	-3	-4	-5	-6
-2	-2	1	0	-1	-2	-3	-4	-2	-3
-3	-3	0	3	2	1	0	-1	-2	-3
-4	-1	-1	2	5	4	3	2	1	0
-5	-2	-2	1	4	7	6	5	4	3
-6	-3	0	0	3	6	6	5	7	6
-7	-4	-1	-1	2	5	8	7	6	9

b) Yes, there is more than one possibility of aligning the given sequences. This happens because when we reach one particular cell while backtracking there can be more than one paths to reach the top left corner of the grid. Hence, we get multiple optimal alignments while the score for each is same.

c) All the optimal alignments are given below. Each one of them have a score of 9.

```
TTC__AGTA
||_|||
ATCAGAGTA

TTCA__GTA
|||_||
ATCAGAGTA

TTCAG__TA
|||_|
ATCAGAGTA
```

Question 2)

a) The generated dynamic programming matrix is:

0	0	0	0	0	0	0	0	0	0
0	0	2	1	0	0	0	0	2	1
0	0	2	1	0	0	0	0	2	1
0	0	1	4	3	2	1	0	1	1
0	2	1	3	6	5	4	3	2	3
0	1	1	2	5	8	7	6	5	4
0	0	3	2	4	7	7	6	8	7
0	2	2	2	4	6	9	8	7	10

b) The best alignments are given below and each one of them have a score of 10 because in local alignment we take the maximum of the cells to be the best score and here 10 is the maximum.

TC__AGTA
TCAGAGTA

TCA__GTA
TCAGAGTA

TCAG__TA
TCAGAGTA

Question 3)

In the local alignment the minimum value that a box can hold in the grid is 0. So, the changes that were required are as follows:

- (1) Initialise row0/col0 with 0."
- (2) Find the cell with the highest value (i, j) and extend the alignment back to the first zero value.
- (3) The score of the alignment is the value in that cell.

Question 4)

The global alignment matrix after change in scoring scheme

0	-2	-4	-6	-8	-10	-12	-14	-16	-18
-2	-1	0	-2	-4	-6	-8	-10	-12	-14
-4	-3	1	-1	-3	-5	-7	-9	-8	-10
-6	-5	-1	3	1	-1	-3	-5	-7	-9
-8	-4	-3	1	5	3	1	-1	-3	-5
-10	-6	-5	-1	3	7	5	3	1	-1
-12	-8	-4	-3	1	5	6	4	5	3
-14	-10	-6	-5	-1	3	7	5	3	7

The local alignment matrix after change in scoring scheme

0	0	0	0	0	0	0	0	0	0
0	0	2	0	0	0	0	0	2	0
0	0	2	1	0	0	0	0	2	1
0	0	0	4	2	0	0	0	0	1
0	2	0	2	6	4	2	0	0	2
0	0	1	0	4	8	6	4	2	0
0	0	2	0	2	6	7	5	6	4
0	2	0	1	2	4	8	6	4	8

The optimal global alignments:

```

TTC__AGTA
||  |||
ATCAGAGTA

TTCA__GTA
|||  |||
ATCAGAGTA

TTCAG__TA
|||  ||
ATCAGAGTA

```

The optimal local alignments

```

TCAG
|||
TCAG

TCAGTA
||| |
TCAG_A

AGTA
|||
AGTA

TCA__GTA
|||  |||
TCAGAGTA

TCAG__TA
|||  ||
TCAGAGTA

```

As can be seen from the results above the change in scoring scheme has changed the result. When aligning biological sequences, the choice of parameter values for the alignment scoring function is critical. Small changes in any score can yield radically different alignments. Since in our problem the score for the gap penalty is changed from -1 to -2 hence there is a difference in the results.