

**Indraprastha Institute of Information Technology Delhi (IIITD)**  
**Department of Computational Biotechnology**

**BIO213 – Introduction to Quantitative Biology**

**ASSIGNMENT-1 (March 18, 2023)**

---

**Question 1.** Compute global alignment between the following DNA sequences using dynamic programming and the given scoring scheme.

DNA sequences: (1) GATGCGCAG, (2) GGCAGTA  
Scoring function: Match = +2, Mismatch = -3, Gap = -1

- a) Provide the bidimensional array obtained for computing the optimal alignment. [5 marks]

		G	A	T	G	C	G	C	A	G
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
G	-1	2	1	0	-1	-2	-3	-4	-5	-6
G	-2	1	0	-1	2	1	0	-1	-2	-3
C	-3	0	-1	-2	1	4	3	2	1	0
A	-4	-1	2	1	0	3	2	1	4	3
G	-5	-2	1	0	3	2	5	4	3	6
T	-6	-3	0	3	2	1	4	3	2	5
A	-7	-4	-1	2	1	0	3	2	5	4

5 marks if all correct.

- b) Is there more than one possibility of optimally aligning the given sequences? [5 marks]

Yes. 5 marks.

- c) If yes, show all the optimal alignments with their scores. If no, provide the best alignment obtained with its corresponding score. [10 marks]

GATGCGCAG--

GATGC-GC-AG

```

| | | |
G--G--CAGTA

```

```

| | | | |
G--GCAG-TA-

```

```

GATGCGCAG--
| | | |
G--GC--AGTA

```

```

GATGC-G-CAG
| | | |
G--GCAGT-A-

```

```

GATGCGCAG--
| | | |
---G-GCAGTA

```

```

GATGCGCAG--
| | | |
G----GCAGTA

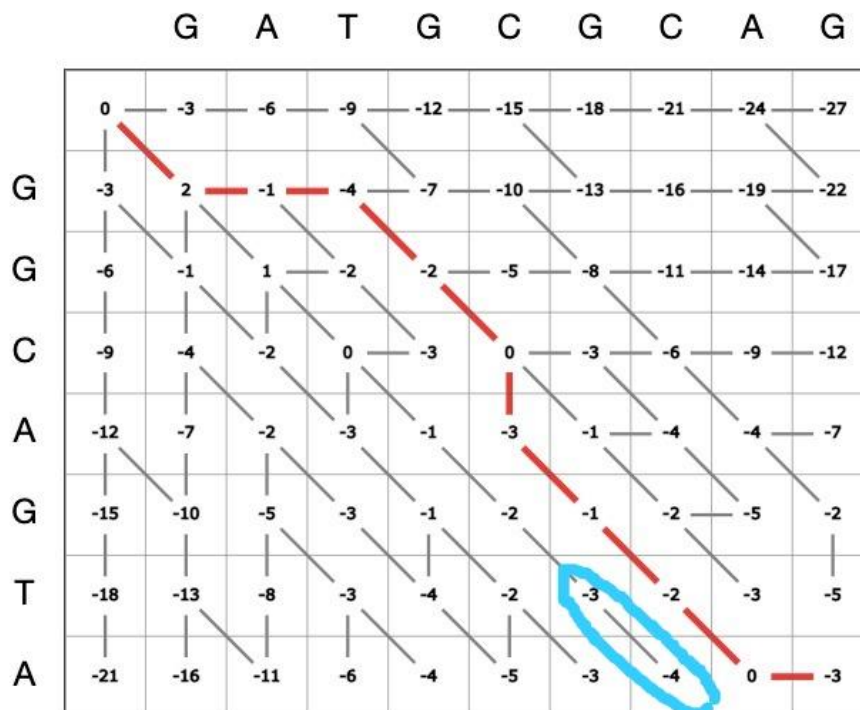
```

1.5 marks for each of the correct alignments, 1 mark extra if all the alignments have been shown.

**Question 2.** Will changing the scoring scheme to [Match = +2, Mismatch = -1, Gap = -3] modify the results obtained in Question 1? If yes, show the results and justify the same. [10 marks]

### Dynamic programming matrix visualisation

Paths for optimal alignments are indicated in red



```

GATGC-GCAG
|  | | | |
G--GCAGTA-

```

5+2 marks for the correct matrix and alignment, respectively.

Now that gaps have more penalty than mismatches, we can see a mismatch in the alignment to get the next matched base-pair. Also, the number of possible alignments has reduced drastically due to the increase in penalty for gaps. 3 marks for the justification.

**Question 3.** Compute the most optimal local alignment (using dynamic programming) for the same DNA sequences, and the scoring scheme provided in Question 2.

a) Provide the generated bidimensional array. [5 marks]

		G	A	T	G	C	G	C	A	G
	0	0	0	0	0	0	0	0	0	0
G	0	2	0	0	2	0	2	0	0	2
G	0	2	1	0	2	1	2	1	0	2
C	0	0	1	0	0	4	1	4	1	0
A	0	0	2	0	0	1	3	1	6	3
G	0	2	0	1	2	0	3	2	3	8
T	0	0	1	2	0	1	0	2	1	5
A	0	0	2	0	1	0	0	0	4	2

5 marks if all correct.

b) Provide all the alignments obtained with their corresponding scores. [10 marks]

```

GCAG
| | | |
GCAG

```

10 marks for correct alignment.

**Question 4.** What changes were required in the program in order to perform local rather than global pairwise sequence alignment? [5 marks]

1. First row and column had to be initialized to zero.
2. Minimum value added to the cell was zero. Negative values not acceptable.
3. Trace back need not start from the last cell of the matrix. It starts with the highest value located anywhere in the matrix.

First point; 1 mark, Second and third point: 2 marks each