# Indian Statistical Institute

M.Tech. (CS), Second Year, Mid-Sem of First Semester Examination, 2025-26

**Computational Molecular Biology and Bioinformatics**

Answer Keys

1. (a) $5^3 * \left(1-\left(1-\frac{1}{5^3}\right)^{10-3+1}\right) \approx 7.8.$

(b) $\frac{1}{1*4*6*6*4} = \frac{1}{576}.$

(c) The complementary strand of a Poly-A tail is ideally shorter than 200nt in length and consists of successive 'T's. Therefore, it cannot include CpG islands.

2. Match = +1, Mismatch = –1, Gap = –2.

a =

| | – | **A** | **A** | **T** | **C** | **G** |
|---|---|---|---|---|---|---|
| – | 0 | 0 | 0 | 0 | 0 | 0 |
| **A** | 0 | 2 | | | | |
| **T** | 0 | | 1 | | | |
| **G** | 0 | | | | | |

b =

| | – | **A** | **A** | **T** | **C** | **G** |
|---|---|---|---|---|---|---|
| – | 0 | –2 | –4 | –6 | –8 | –10 |
| **A** | –2 | 2 | | | | |
| **T** | | | | | | |
| **G** | | | | | | |

c =

| | – | **A** | **A** | **T** | **C** | **G** |
|---|---|---|---|---|---|---|
| – | 0 | –2 | | | | |
| **A** | –2 | 2 | | | | |
| **T** | –4 | | | | | |
| **G** | –6 | | | | | |

The final alignment is as follows.

AATCG
– ATG –

3. (a)

**Sequence:** CAA

**Sequence:** TG CA AA
**Hashcode:** 6 16

| | | | | | | |
|---|---|---|---|---|---|---|
| | ↖ | ↖ | ↗ | ↗ | | |
| **Sequence:** | T | G | C | A | A | A |
| **Position:** | 1 | 2 | 3 | 4 | 5 | 6 |
| **Hashcode:** | 1 | 2 | 3 | 4 | 4 | 4 |

| Hashcode | Position | # Positions | Sequence/x-mer |
|---|---|---|---|
| 1 | 1 | 1 | T |
| 2 | 2 | 1 | G |
| 3 | 3 | 1 | C |
| 4 | 4, 5, 6 | 3 | A |
| 6 | 1 | 1 | TG |
| 16 | 3 | 1 | CA |
| … | … | … | … |

(b) The lazy Needleman-Wunsch algorithm helps to identify the optimal alignment overlapping a given offset, thereby figuring out the indels.

4. (a) We can find out dependencies between different datasets attributed to a particular cancer by assuming there are separate sets of random variables generating a particular dataset.

(b) Given a directed network, we can send 2 units of flow from source to sink. Each arc has a linear per-unit cost and a capacity. Additionally, opening an arc (using it at all) incurs a fixed setup cost. Suppose you want the minimum-cost way to send 2 units. This can be formulated as an MILP by figuring out the objective function (minimize), capacity/linking constraints, flow conservation, and other feasibility constraints.