CSC 314**, Bioinformatics Lab #10: Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Pairwise Sequence Alignments**

Consider the alignments below of the sequences

QNISEKDLLDTENK and QSEKDAADTKEVV

**Alignment 1**

1 QNISEKDLLDTENK- 14

.||||..||:..

1 --QSEKDAADTKEVV 13

**Alignment 2**

1 QNISEKDLLDTENK--- 14

| ||||..|| |

1 Q--SEKDAADT--KEVV 13

For each alignment, use the BLOSUM-62 matrix (available in your notes) and the gap opening and extension penalties below to calculate a *semi-global alignment* score. **Note: in a semi-global alignment, gaps at the beginning or end of an alignment are not penalized (i.e., no gaps are penalized in Alignment 1; only the gaps in the second sequence are penalized in Alignment 2).**

1. BLOSUM-62 substitution matrix, Gap Opening Penalty of 10, Gap Extension Penalty of 1.
   1. Fill in the bottom row of the table below to specify the score for each pair of the alignment for Alignment 1:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Q | N | I | S | E | K | D | L | L | D | T | E | N | K | - |
|  |  | . | | | | | | | | | . | . | | | | | .. | . | . |  |
| - | - | Q | S | E | K | D | A | A | D | T | K | E | V | V |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

* 1. Total score for alignment 1 =
  2. Fill in the bottom row of the table below to specify the score for each pair of the alignment for Alignment 2:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Q | N | I | S | E | K | D | L | L | D | T | E | N | K | - | - | - |
| | |  |  | | | | | | | | | . | . | | | | |  |  | | |  |  |  |
| Q | - | - | S | E | K | D | A | A | D | T | - | - | K | E | V | V |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

* 1. Total score for alignment 2 =
  2. Of the two alignments, which alignment is more optimal?

1. BLOSUM-62 substitution matrix, Gap Opening Penalty of 1, Gap Extension Penalty of 0.5.
   1. Fill in the bottom row of the table below to specify the score for each pair of the alignment for Alignment 1:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Q | N | I | S | E | K | D | L | L | D | T | E | N | K | - |
|  |  | . | | | | | | | | | . | . | | | | | .. | . | . |  |
| - | - | Q | S | E | K | D | A | A | D | T | K | E | V | V |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

* 1. Total score for alignment 1 =
  2. Fill in the bottom row of the table below to specify the score for each pair of the alignment for Alignment 2:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Q | N | I | S | E | K | D | L | L | D | T | E | N | K | - | - | - |
| | |  |  | | | | | | | | | . | . | | | | |  |  | | |  |  |  |
| Q | - | - | S | E | K | D | A | A | D | T | - | - | K | E | V | V |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

* 1. Total score for alignment 2 =
  2. Of the two alignments, which alignment is more optimal?

1. What is the percent identity for
   1. Alignment 1
   2. Alignment 2?
2. What is the percent similarity for
   1. Alignment 1
   2. Alignment 2

Note: In the alignment, a single (.) or double (..) dot denotes similarity

Consider the possible scoring systems and then answer the questions below.

|  |  |  |
| --- | --- | --- |
| Scoring System | Substitution Matrix | Alignment type |
| #1 | BLOSUM90 | Local alignment |
| #2 | BLOSUM45 | Local alignment |
| #3 | BLOSUM90 | Semi-global alignment |
| #4 | BLOSUM45 | Semi-global alignment |

1. Which scoring system above is most appropriate for an alignment involving
2. two homologous proteins that are expected to have a very high *overall* similarity (the proteins are nearly identical).
3. two homologous proteins that are distantly related and may have a relatively low similarity (due primarily to missense mutations)
4. two different proteins where the goal is to identify *small functional regions* (domains) of very high similarity.