

CSC 314, Bioinformatics Lab #8:
Pairwise Sequence Alignments

Name: _____

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.
 - a. Score for alignment 1 =
 - b. Score for alignment 2 =
 - c. Of the two alignments, which alignment is optimal?
2. BLOSUM-62 substitution matrix, Gap Opening Penalty of 1, Gap Extension Penalty of 0.5.
 - a. Score for alignment 1 =
 - b. Score for alignment 2 =
 - c. Of the two alignments, which alignment is optimal?

3. What is the percent identity for (a) Alignment 1 and (b) Alignment 2?

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

Scoring System	Substitution Matrix	Gap Opening Penalty	Gap Extension Penalty
#1	BLOSUM90	2	1
#2	BLOSUM45	2	1
#3	BLOSUM90	10	1
#4	BLOSUM45	10	1

4. Which scoring system above is most appropriate for an alignment involving
- a) two homologous proteins that are expected to have a very high overall similarity (the proteins are nearly identical and have few, if any, gaps).
 - b) two homologous proteins that are distantly related and may have a relatively low similarity (due primarily to missense mutations, not gaps)
 - c) two different proteins where the goal is to identify small functional regions (domains) of very high similarity.