CSC 314, Bioinformatics Lab #9:   **Name:**_____
Pairwise Sequence Alignments

Consider the alignments below of the sequences

QNISAKDLLDTENK and QSEKDALATKEVV

Alignment 1

```
QNISAKDLLDTENK-
  .|.||.|.|:..
--QSEKDALATKEVV
```

Alignment 2

```
QNISAKDLLDTENK---
|   |.||.|.|   |
Q--SEKDALAT--KEVV
```

For each alignment, use the BLOSUM-62 matrix from 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. Fill in the bottom row of the table below to specify the score for each position
      of the alignment for Alignment 1:

Q	N	I	S	A	K	D	L	L	D	T	E	N	K	-
		.	|	.	|	|	.	|	.	|	..	.	.	
-	-	Q	S	E	K	D	A	L	A	T	K	E	V	V
0	0	-3												0

   b. Total score for alignment 1 =

c. Fill in the bottom row of the table below to specify the score for each position of the alignment for Alignment 2:

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

d. Total score for alignment 2 =

e. Of the two alignments, which alignment is more optimal?

2. BLOSUM-62 substitution matrix, Gap Opening Penalty of 1, Gap Extension Penalty of 0.5.

   a. Fill in the bottom row of the table below to specify the score for each position of the alignment for Alignment 1:

Q	N	I	S	A	K	D	L	L	D	T	E	N	K	–
		.		.			.		.		. .	.	.	
–	–	Q	S	E	K	D	A	L	A	T	K	E	V	V
0	0	–3												0

b. Total score for alignment 1 =

c. Fill in the bottom row of the table below to specify the score for each position of the alignment for Alignment 2:

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

d. Total score for alignment 2 =

e. Of the two alignments, which alignment is more optimal?

3. Note that in these alignments, a vertical bar denotes identity, and a double dot (..) denotes similarity, and a single dot (.) denotes a mismatch.

   a. What is the percent identity for Alignment 1?

   b. What is the percent similarity for Alignment 1?

   c. What is the percent identity and percent similarity for alignment 2 (note that these are the same because there are no similar amino acid pairs in the alignment)?

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

Scoring System	Substitution Matrix	Gap opening penalty
#1	BLOSUM90	1
#2	BLOSUM90	10
#3	BLOSUM45	1
#4	BLOSUM45	10

Which scoring system above is most appropriate for an optimal alignment involving

   a) two homologous proteins that are expected to have a very high overall similarity (the proteins are nearly identical), and few, if any gaps. Note: to limit the number of gaps in an alignment, you want the gap opening penalty to be *high*.

   b) two proteins that are distantly related, and are expected to differ because of a large number of missense mutations, but few, if any insertions and deletions.

5. Create a dot plot for the two sequences EASTERN and STATES. Based on your dot plot, what are the two local areas of similarity (matching sequences of 2 or more characters)?

	S	T	A	T	E	S
E						
A						
S						
T						
E						
R						
N						