1. (10 points) We are going to trace the following DNA sequence through the process of transcription and translation:

TTCATGACTAAGCCC

(a) (2 points) Write down the RNA version of the DNA sequence.
(b) (2 points) Write down the amino acid sequence corresponding to the RNA sequence.
(c) (2 points) Write down the reverse complement of the DNA sequence.
(d) (2 points) Write down the RNA version of the reverse complemented DNA sequence.
(e) (2 points) Write down the amino acid sequence of the reverse complemented RNA sequence.

2. (20 points)

A R N D C Q E G H I L K M F P S T W Y V B Z X

A
-2 6 -1 -2 -4 -1 1 0 -3 -3 -2 -1 -1 -3 -2 0 -1 -1 -1 -1 -1 -1
B -2 -6 1 -2 -4 -1 0 -3 -2 -1 -1 -3 -2 0 -1 -1 -1 -1 -1 -1 -1
C -2 -1 6 -4 -1 1 2 -2 -4 -5 -1 -4 -4 -2 -1 -1 -6 -4 -4 4 1 -2
D C -1 -4 -3 -4 9 -4 -5 -4 -4 -2 -2 -4 -2 -3 -2 -1 -3 -3 -1 -4 -4 -3
E Q -1 1 0 -1 -4 6 2 -2 -1 -3 -1 1 0 -4 -2 0 -1 -3 -2 -3 0 -3 -1
F E -1 -1 1 1 -5 2 -6 -3 0 -4 -4 1 -2 -4 -2 0 -1 -4 -3 -3 1 4 -1
G R 0 -3 -1 -1 -2 -4 -3 6 -3 -5 -4 -2 -4 -3 -1 -2 -4 -4 -1 -3 -2
H 0 -2 0 -2 -4 1 0 -3 8 -4 -3 -1 -2 -3 -1 -2 -3 2 -4 -1 0 -2
I L -2 -3 -4 -4 -2 -3 -4 -5 -4 5 1 -3 1 -1 -4 -3 -1 -3 -2 3 -4 -4 -2
J T -2 -3 -4 -5 -2 -3 -4 -4 -3 1 4 -3 2 0 -3 -3 -2 -2 2 1 -4 -3 -2
K G 1 2 0 -1 -4 1 1 2 -1 -3 -3 5 -2 -4 -1 -1 -1 -4 -3 -3 1 1 -1
L S 0 -1 -2 -3 -4 -2 0 -2 -4 -2 1 2 -2 6 0 -3 -2 -1 -2 2 1 -3 -2 -1
M F -3 -4 -4 -4 -3 -4 -4 -4 -2 -1 -0 -4 0 -6 -4 -3 -2 0 3 -1 -4 -2
N P -1 -2 -3 -2 -4 -2 -3 -3 -4 -3 -1 -3 -4 -8 -1 -2 -4 -5 -4 -3 -2 -2
O S 1 1 -1 -1 -2 0 0 -1 -1 -1 -3 -1 -2 -3 -1 5 1 -4 -2 -2 0 0 -1
P T 0 -1 0 -1 -1 1 -1 -2 -2 -1 -2 -1 -1 -2 -2 1 5 -4 -2 0 -1 -1 -1
Q W -3 -4 -4 -6 -3 -3 -4 -4 -3 -3 -2 -4 -2 -0 -5 -4 -4 11 2 -3 -5 -4 -3
R Y -2 -3 -4 -3 -4 -3 -4 -3 -2 -3 -3 -4 -2 -3 3 -4 -2 -2 2 -7 -2 -3 -3 -2
S V 0 -3 -4 -4 -1 -3 -3 -4 -4 -3 1 -3 1 -1 -3 -2 0 -3 -2 4 -4 -3 -1
T B -2 -2 4 -4 -4 -0 1 -1 -1 -1 -1 -1 -1 -2 -2 0 -1 -5 -3 -4 -4 0 -2
U Z -1 0 0 1 -4 3 -4 -3 0 -4 -3 1 -2 -4 -2 0 -1 -4 -3 -3 0 4 -1
V X -1 -1 -1 -2 -3 -1 -1 -2 -2 -2 -2 -1 -1 -2 -2 -1 -1 -3 -2 -1 -2 -1 -1

GGL--RPSKLDPV--YSWWVLSSLAIGRLD
GRLNGRP--KLPDVC--YSQWVLAS---GRLH

(a) (10 points) Compute the score of the above alignment using the BLOSUM80 substitution matrix given above and a linear gap penalty of -5.
(b) (10 points) Compute the score of the above alignment using the substitution matrix given above and an affine gap penalty of -5 to open a gap and -1 to extend a gap.

3. (30 points)

(a) (15 points) Draw and fill in the global alignment dynamic programming matrix for these two sequences: “MLHCA” and “HNCSV.” Use the BLOSUM80 matrix and a linear gap penalty of -4.

Be sure to include traceback arrows. Report the resulting global alignment as well as its score.
(b) (15 points) Repeat the previous question, but use the local alignment algorithm.

4. (40 points) Write a Python program to score a given alignment using an affine gap penalty. The program should take as input the following arguments: the name of a file containing the substitution matrix in the three-column format we discussed in class, the name of a file containing the two sequences (including gap characters), the gap open penalty and the gap extension penalty.

Points will be assigned as follows:
(a) (6 points) Read the inputs correctly from the command line.
(b) (10 points) Read the substitution matrix into a dictionary.
(c) (2 points) Check that the sequences are the same length, and report an error if they are not.
(d) (5 points) Loop over the sequences in parallel.
(e) (5 points) Compute substitution scores for aligned characters.
(f) (10 points) Compute gap penalties correctly.
(g) (2 points) Print error messages to stderr and the final score to stdout.

You can test your program on the alignment in Question 2. You should email the code to the tutors, who will test that it runs correctly on a different alignment.