Practice Questions
(Answers to the questions are enclosed with blue rectangles)

Question 1

Based on the completed dynamic programming alignment matrix below, answer the following questions:

a. (2 point) Is this a Needleman-Wunsch (global) or Smith-Waterman (local) alignment?

Smith-Waterman (local)

b. (2 points) What was the nucleotide match score parameter? Assume that any match (e.g., A-A, C-C, G-G, or T-T) receives the same score.

5

c. (2 points) What was the mismatch penalty? Assume that any mismatch receives the same penalty. This should be a negative number.

-2

d. (2 points) What was the gap penalty? Assume we used a fixed (linear) gap penalty. This should also be a negative number.

-3

e. (2 points) What is the score of the highest-scoring local alignment and is the alignment unique?

17 and it is unique

		T	C	T	C	A	C
	0	0	0	0	0	0	0
T	0	5	2	5	2	0	0
C	0	2	10	7	10	7	5
C	0	0	7	8	12	9	12
A	0	0	4	5	9	17	14

Question 2

Let $s(i, j)$ be the log-odds score of the amino-acid pair (i, j).

(a) What is the interpretation of a negative value of $s(i, j)$?

Answer. A negative value of $s(i, j)$ means that the observed frequency of exchanging i and j is less than the expected frequency. This substitution is disfavored.

(b) What is the interpretation of $s(i, j) = 0$?

Answer. A value of 0 for $s(i, j)$ means that the observed frequency of exchanging i and j is the same as the expected frequency. The substitution of i for j is neutral.

(c) What is the interpretation of a positive value of $s(i, j)$?

Answer. A positive value for $s(i, j)$ means that the observed frequency of exchanging i and j is greater than the expected frequency. This substitution is potentially favorable.

Question 3

(a) Fill out the dynamic programming table for determining the optimum global alignment between sequences ACTG and CGGA. Assume that a match is scored $+3$ and that mismatches and spaces are scored -1 each.

		C	G	G	A
	0	-1	-2	-3	-4
A	-1	-1	-2	-3	0
C	-2	2	1	0	-1
T	-3	1	1	0	-1
G	-4	0	4	4	3

(b) What is the optimum alignment corresponding to the table in part (a) and what is its score?

Answer.

-	C	G	G	A
A	C	T	G	-
-1	3	-1	3	-1

The score of this alignment is 3.

Question 4

Consider the following multiple alignment of DNA sequences

	1	2	3	4	5
S_1	-	G	A	G	C
S_2	C	T	A	G	A
S_3	C	G	A	-	A
S_4	A	G	C	G	A

(a) Give the (average) profile of the above alignment by filling out the table below.

	1	2	3	4	5
A	0.25	0	0.75	0	0.75
C	0.5	0	0.25	0	0.25
T	0	0.25	0	0	0
G	0	0.75	0	0.75	0
-	0.25	0	0	0.25	0

(b) (6 points) Suppose we use match score 1.86, mismatch score -3 (PAM 10), and gaps are scored -2. What is the score of aligning a nucleotide G with column 2 of the profile? What is the score of aligning nucleotide G with column 4?

G with column 2:

$S(G,T) * 0.25 + S(G,G) * 0.75$

$= -3 * 0.25 + 1.86 * 0.75$

$= -0.75 + 1.395$

$=0.645$

G with column 4:

$S(G,G) * 0.75 + S(G,-) * 0.25$

$=1.86 * 0.75 + -2 * 0.25$

$=1.395 - 0.5$

$=0.895$