Seven11 Assignment 1                      Due 1:30 pm, Friday, September 21

Your name:

This assignment is required for students taking the 12-unit course only. In class, we discussed pairwise sequence alignment with a linear gap penalty, where the cost of an alignment with indels is independent of the placement of the indels within the alignment. In other words, the penalty for \( k \) adjacent indels is the same as the penalty for \( k \) indels at isolated positions within the alignment.

The linear gap penalty does not reflect the fact that the gain or loss of an entire sequence fragment in a single event is less likely to disrupt protein structure than the gain or loss of individual residues spread across disparate locations. Alignments with a small number of large gaps are more readily obtained with a gap penalty that includes a high cost for the first indel and a much lower cost for adding an indel to an existing gap. In this assignment, you will use an affine gap penalty function that has this property, described in detail in:

  https://ereserves.library.cmu.edu/reserves/esdurand2.pdf

The first question on this assignment asks you to design a global pairwise alignment that penalizes leading and internal gaps differently. The second and third questions ask you to align the same pair of sequences with a linear gap function and with an affine gap function, for comparison. In both cases, leading and internal gaps will be penalized differently. Turn in your handwritten answers on the alignment templates that are available on the website. You may not use a program to do this homework.

Collaboration is allowed on this homework. You must hand in homeworks individually and list the names of the people you worked with.
1. Suppose you wish to calculate the global alignment of a pair of sequences, $s$ and $t$, with different penalties for leading gaps (i.e., all gaps that appear before the first symbol in $s$ or the first symbol in $t$) and internal gaps (i.e., gaps that appear after the first symbol in $s$ or $t$).

Modify the global pairwise alignment algorithm to achieve this with the following scoring system. Matches: $M = 4$, mismatches: $m = -2$, leading gaps: $g_L = -6$, internal gaps: $g_I = -4$.

For each of the following components of the pairwise global alignment algorithm (pp. 7-8 on the class notes), either give the expression(s) to modify the algorithm to account for different leading and internal gap penalties or write “no change”.

(a) Initiation

(b) Recursion

(c) The cell containing the optimal alignment score

(d) Trace back
2. Alignment with linear gap penalties:

(a) Compute the global alignment of “HIKER” with “MIMICKED”, using the modified pairwise global alignment algorithm from Problem 1, with the scoring system. Matches: $M = 4$, mismatches: $m = -2$, leading gaps: $g_L = -6$, internal gaps: $g_I = -4$. Show your alignment matrix with scores and traceback.

(b) What is the score of the optimal alignment? How many different optimal alignments are there? Show them.

3. Alignment with affine gap penalties:

(a) Recompute the global alignment of “HIKER” with “MIMICKED”, with an affine gap function with match score of 4 ($M = 4$), a mismatch penalty of -2 ($m = -2$), and a gap extension penalty of -1 ($g = -1$). The gap opening penalty is of -6 for leading gaps ($h_L = -6$) and -4 for internal gaps ($h_I = -4$). Recall that cost of the first gap in a sequences of indels is $h + g$ and the remaining gaps cost $g$. Show all three alignment matrices.

(b) What is the score of the optimal alignment? How many different optimal alignments are there? Show them.