

Problem

Consider the following substitution matrix for DNA sequences:

	A	C	G	T
A	10	2	5	2
C	2	10	2	5
G	5	2	10	2
T	2	5	2	10

Question 1: What is the optimal (here maximal) cost of an alignment of AATAAT and AAGG using the above substitution matrix and gap cost -5?

Question 2: What is the optimal (here maximal) cost of an alignment of seq1.fasta and seq2.fasta using the same substitution matrix and gap cost? (You probably want to implement the algorithm for computing the cost of an optimal alignment.)

Question 3 (optional): How does an optimal alignment look like for the above two pairs of sequences using the given substitution matrix and gap cost -5? (you probably want to implement the algorithm for finding an optimal alignment by backtracking through the dynamic programming table.)

Question 4 (optional): How many optimal alignments are for the above two pairs of sequences using the given substitution matrix and gap cost -5? Explain how you can compute the number of optimal alignments.

To test your programs you are welcome to use the two test examples in project1_examples.txt