

Assignment 1.

Problem 1. Sequence Alignment.

20 points.

Calculate and show the Dynamic Programming matrix and an optimal alignment for the DNA sequences GCATTGC and GATTAGC, scoring +2 for a match, -1 for a mismatch, and a gap penalty of 2 (i.e., each gap column contributes -2). (If there are more than one optimal alignments, you can report any of them.)

Problem 2. Pairwise Sequence Alignment.

20 points.

Find the optimal pairwise global alignment of the sequences TCTGC**C**TCTGC and ACTGAC**C**ACTGAC with the condition that the **C** nucleotides shown in bold font must be aligned with each other. What is the optimal alignment score? The scoring parameters are defined as +2 for match, -1 for mismatch and -2 for gap. Show the dynamic programming matrix you used. (And please do not create a full 11x13 matrix.)

Problem 3. (Dynamic programming)
Points: 30.

(A) 10 points.

Consider the task of finding the optimal global alignment of the following two sequences: ATC and ATATCTC.

The scoring function is:

MATCH : 5.

MISMATCH : -10.

GAP : -5.

Construct the dynamic programming table and write down the optimal alignment.

(B) 10 points.

Consider the task of finding the optimal local alignment of the following two sequences: ATC and ATATCTC.

The scoring function is:

MATCH : 5.

MISMATCH :-10.

GAP : -5.

Construct the dynamic programming table and write down the optimal alignment.

(C) 10 points.

According to (A) and (B), which one is more suitable for aligning these two sequences? Why?

Problem 4 (bonus). Sequence alignment.

30 points.

Consider the affine gap penalty:

gap_start = the cost of starting a gap

gap_extend = the cost of extending a gap by one more space

1. Write down a dynamic programming algorithm for global pairwise alignment with the affine gap penalty. You can assume that the match score and mismatch score are also given.
2. Provide the complexity of your algorithm. Is it optimal?