

Problem 1. Profile alignment

In multiple sequence alignment, we often have to align a profile to a sequence. Use the average Sum of Pairs score to align the given sequence and profile. Your answer should include the filled in matrix given below, arrows for the traceback procedure and the final alignment.

The sum of pairs score, $S(m_i)$ for the i^{th} column is defined as

$$S(m_i) = \sum_{k < l} s(m_i^k, m_i^l)$$

where k and l iterate over all rows of the alignment. Use the following scoring criteria $s(x, -) = 0$, $s(-, -) = 0$ where $-$ denotes a gap. $s(x, y) = 2$ when x matches y , and $s(x, y) = 1$ when x and y don't match.

		A	G	G	T
		A	-	G	T
A					
G					
A					

Problem 2. Read the paper:

Evaluating the accuracy and efficiency of multiple sequence alignment methods.

- When comparing MSA algorithms, what are the main factors that should be considered?
- How to measure the accuracy of a given MSA algorithm?
- Select two algorithms mentioned in the paper. Compare and contrast them by giving most distinguishing characteristics for each.
- Use the online tools of algorithms in part c) to perform MSA for the following proteins:

COX2_BETVU
COX2_BREAN
COX2_BRECS
COX2_BRENA
COX2_CHOCH
COX2_CYACA
COX2_DAUCA

Show the alignment results and matrix.