

Bioinformatics

Exercises - Sequence alignment

1. Given the following two sequences:

HEAGAWGHEE
PAWHEAE

- a) Build the alignment matrix for these sequences and compute the optimal global alignment using the Needleman-Wunsch algorithm (use a gap penalty value of -8 and the BLOSUM50 scoring matrix)
 - b) Build the alignment matrix for the two sequences, and compute the optimal local alignment using the Smith-Waterman local alignment algorithm.
2. Using the Randseq tool (<http://expasy.org/tools/randseq.html>), build two sequences AB and BA, where A and B are two random aminoacid sequences of length 50. Produce the local and global alignment of the AB and BA sequences using the "needle" and "water" options of the Align tool (<http://www.ebi.ac.uk/emboss/align>). How do you interpret the obtained results? Now use the LALIGN tool (http://www.ch.embnet.org/software/LALIGN_form.html) to list sub-optimal alignments of the AB and BA sequences. Compare with the previously obtained result. Do you think this example has any biological relevance? Why?
 3. Align the sequences of the human coagulation factors IX and XII locally and globally. Use the LALIGN program to list the sub-optimal alignments. Compare the resulting alignments with the Uniprot domain definitions for the two protein sequences.