EE550
Computational Biology
Homework 3
Due 12.04.2022

Instructions: Please return your homework as a single PDF document by the due date with "EE550" and your student number in the file name. Sharing of ideas and discussions is encouraged but sharing of results and/or text is not. Show the details your work including the intermediary results.

Question (100 points) Viral miniproteins are proteins of viral origins that are typically shorter than 50 amino acids long (see DiMaio, Daniel. "Viral miniproteins." Annual review of microbiology vol. 68 (2014): 21-43. doi:10.1146/annurev-micro-091313-103727). For this homework, you will need to identify two homologous but non-identical miniproteins of your choosing from the UniProt Knowledgebase (at the url https://www.uniprot.org/) and carry out a pairwise alignment between them using dynamic programming.

a) (20 points) Download the amino acid sequences of the two proteins and provide brief descriptions for both.
b) (60 points) Carry out a pairwise alignment of the amino acid sequences of the two proteins using the original dynamic programming alignment strategy or an improved version that uses a log-odds scoring matrix instead of α, β and a suitable choice for the ω parameter. Make sure that you present the intermediary result of a properly filled alignment matrix A with the marking of the optimal alignment along with the final alignment result with the matches and gaps.
c) (20 points) Align the two proteins using BLAST and compare the two alignment results.