Instructions

This is a take-home test. Turn in your answers in a single PDF document via e-mail. Correct answers to wrong questions are not creditable. Verbal discussions and idea sharing on the questions are encouraged, but sharing text and/or code and/or results in any manner, shape or form is absolutely not and should be avoided. Your name on the answer document acts as a statement from your part that you have neither given nor received any unauthorized help. Show your work.

For the questions below, consider the following list of proteins with UniProt identifiers

P10360, P56423, P02340, P10361, Q9WUR6, Q8SPZ3, P41685, P61260, O36006, P13481, P79734, P56424, Q9TUB2, P51664, P67939, Q95330, Q9TTA1, P67938, Q29537, O09185, P79892, P04637

Question 1 (20 points) Carry out a multiple sequence alignment on the proteins listed above via progressive sequence alignment by dynamic programming with your choice of parameters \(\alpha\), \(\beta\), and \(\omega\).

(Hint: Use average similarity when evaluating the similarity between the sites of two sequence clusters.)

Question 2 (20 points) Using the multiple sequence alignment results obtained above, calculate suitable scoring matrices using PAM and BLOSUM strategies.

(Hint: Use all non-gapped aligned sites when processing each protein pair and adjust the calculation parameters accordingly.)

Question 3 (20 points) Re-align the sequences using the scoring matrices you obtained above to evaluate site similarity instead of the original choices of \(\alpha\) and \(\beta\) parameters.

Question 4 (20 points) Re-calculate the PAM and BLOSUM scoring matrices using the new alignments.

Question 5 (20 points) Compare and contrast the scoring matrices obtained following the two different alignments to each other as well as the established PAM and BLOSUM scoring matrices.