EE550 Computational Biology
Spring 2022 Midterm
Due 25.4.2022 midnight

Instructions

This is a take-home test. Turn in your answers in a single PDF document with your name and student identification number in the file name. 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 this test, consider the proteins assigned to you as provided in the table below.

Student ID	UniProt ID	Student ID	UniProt ID	Student ID	UniProt ID
293088009	Q9Y243	293088010	P15336	293085009	Q9H2K8
293088002	P42574	293095016	P19838	283085006	P19438
293088003	Q16539	283085045	Q13164	293085014	P45983
293088006	P11831	283097043	Q13163		

Question 1 (10 points) Conduct a literature search on your assigned protein and provide a summary description on its involvement in various cellular processes.

Question 2 (30 points) Carry out sequence alignment of your protein in an online protein database of your choosing and identify at most 20 proteins similar to yours, excluding isoforms. Show the alignment statistics as well as a graphical display of the resulting alignments.

Question 3 (30 points) Carry out a multiple sequence alignment over the proteins identified in Question 2 along with the original protein. As usual, show the alignment statistics as well as a graphical display of the resulting alignments and all relevant data.

Question 4 (30 points) Using the non-gapped matches between the sequence pairs in the alignment above, derive PAM and BLOSUM log-odds scoring matrices and compare the most/least preserved amino acids as well as the amino acid pairs that are most/least likely to substitute for each other.