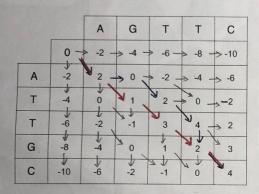
ne: Tia	nru Zhang PID: A15432
	Q1. Why does the DNA sequence have more dots than the protein sequence plot? HINT: what do you know about DNA composition vs protein composition?
	Because DNA only has 4 nucleotides(characters), but protein has 20 amino acids(20 characters), there is higher chance for 2 random DNA sequences to match compared to protein sequences. There are more signal in protein compared to DNA.
	Q2. How can we increase the signal to noise ratio?
	(more noise in DNA sequences compared to protein) Increase window size(the number of monomers included in the sequence) and decrease the stringency moderately to have more signals.
	Q3. What does a 'Match stringency' larger than 'Window size' yield and why?
	It will give an error, because the match stringency specifies number of match characters required per window, and cannot be larger than window size(number of total characters included). For example, when the window size is 4, the number of maximum match is 4, so the stringency cannot be more than 4.
:h'	are the major weaknesses of this approach? HINT: is your inner nerd happy with this? How would you use it to determine if a second set of sequences was more similar to r than a first set of sequences?
	nly visual/graphical and qualitative representations, no numerical/quantatative representations ner two sequences are more related. Inefficient when comparing long sequences.

Section 2: Needleman-Wunsch Alignment

Sequence alignment methods often use something called a 'dynamic programming' algorithm that can be usefully considered as an extension of the dot plot approach. Here we have two sample sequences, and we'd like to use the Needleman-Wunsch algorithm discussed in class to align them. Alignment 2:

Sequence 1: ATTGC Sequence 2: AGTTC Alignment: A_TIGC
AGITC
AGITC
AGTTC
AGTTC
ATTGC
AGTTC
ATTGC
AGTTC
ATTGC



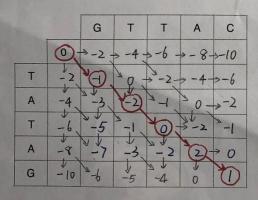
Q5. Using a match score of +2, a mismatch score of -1, and a gap score of -2. Fill in the table below and translate it into a alignment. What is the optimal score for this alignment? Is there one unique alignment with this score?

> optimal Alignment store: 4 There are 2 alignments.

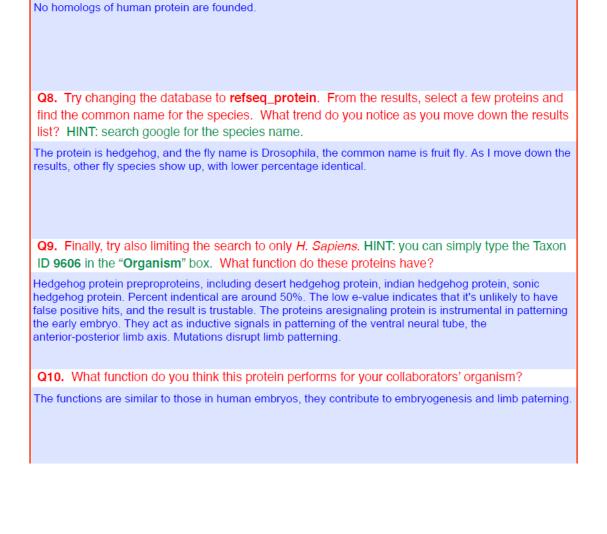
Practice makes perfect. Again use the Needleman-Wunsch algorithm discussed in class to align the following sequences:

Sequence 1: TATAG Sequence 2: GTTAC

Optimal Alignment Score: 1



Q6. Using a match score of 2, a mismatch score of -1, and a gap score of -2. Write out your alignment matrix (table), fill in the values and translate your results into all optimal alignments. What is the optimal alignment score for these sequences? Write out all alignments consistent with this score?



Q7. Using the default settings for NCBI BLAST, can you find any homologs for this protein in Humans? HINT: try using the *LIMITS* and *FILTERING* options we covered in the last lab.