Homework 2: Sequences alignment

Question 1:

Compute by hand an optimal global (using Needleman-Wunsch) and local (using Smith-Waterman) alignment of the sequences ‘NEEDLEMAN’ and ‘DEMLENAEN’. Use as scoring scheme +1 for a match and -1 otherwise.

Question 2:

Use R (see function pairwiseAlignment in the Biostrings package) to compute a local alignment (Smith-Waterman) and a global alignment (Needleman-Wunsch) between the 2 following sequences:

>seq1
AATGCTGTACGTACGCTAGCTA
>seq2
AATGCCGTACGCACACTAGCTCA

What can you observe ? Why ?

Re-do the same operations for the following sequences:

>seq1
GGCATCTCGGTGTGTTCTTGCCCGTGCTGATGTTTGACCCTTGAGCAGAGGTTGCCCTTG
GTGAATTCGCTTTATTTATGTTGAATCACACAAAGGCAACTTTTGTTTGAGTATCAAATCC
TGCTTGGGATGGCTTCCGGGACCCAGTGGCAAGCTCAGGGGCATCTACACCCCTCCCGT
GAGCAAGAATGGACGGGGTAG
>seq2
GCCTGGATGTCTGTACAAGTCTCTGGCCACCTGTAGGCAGGAGAGAACTCTGCAAGTCA
GTCCATTGCTAGTCTGCTTTGCAGTTGGACCTGGGGGCTCATCGGAGGTAATGCCAGCCT
CCCTAGAGGGTGGGAGTCAGGGAGCCAGAGCCGAGGGAGGCAGGGCTGCCTGGGGAA
GAGGCGCCCTGCTCAACTCTCGGCCTGTGCTGACGTTTGATTCTTGAGCAGAGGTTGCC
CTTGG TGAATTCGCTT

In real life, can we use these tools ? Why ?

Question 3:

Go to the NCBI web site and search for the following protein: **NP_188876**

- What is the protein name ?
- From which organism is it from ?
- What is its sequence (in FASTA format) ?

Go to the FASTA algorithm homepage and perform a protein:protein alignment with this sequence against the swissprot database.

Go to the BLAST homepage and perform a protein BLAST with this sequence against the swissprot database.

Can you observe differences between the FASTA and BLAST results? If so, how can you explain them?

Select all the proteins resulting of the BLAST query and download the file with all their FASTA sequences.

Go to the Clustalw homepage and perform a multiple sequence alignment with the default options using this file as input.

What can you observe ? What does it mean from a biological point of view ? Can you make a link between your observations and what we know about the protein we used at the beginning ?

Re-do the Clustalw operation using "id" as scoring matrix. What can you observe ? Why ? Use the help of clustalw to know the difference between the scoring matrix.