Programming EC 1- Alignments DNA + Protein alignment

Packages:

- Numpy
- bioPython (Bio)
- sys

This program runs by inputting two file names (nucleotide, protein) like so:

Python seqAll.py quagganuc.txt quaggapro.txt

(sequence files must be in the same folder as the program)

Output will show all local alignments with the same maximum score, then all global alignments with the same maximum score

This program uses the biopython library to translate the dna sequence to a protein sequence, then preforms alignments. It uses the blosum62 scoring matrix and an affine gap penalty of -10 for opening a gap and -1 for extending a gap. The alignment program uses the Smith-Waterman algorithm to score the local alignment and the Needleman-Wunsch algorithm to score the global alignment.

Examples:
Myoglobin Isoform 1
Global:
FEQULOPINITY PEPTLEF. SSSCANGLE DOES QUILLE FROM ANY VERY SECTION OF THE SECTION
-MAIS DEBIQUILINWAKWEAD FORGGEVILILFRENFETLEFEK WHILESEDBRASEDLER HEAVULALGGILERHESHEAS INPLAGRATHRI PWYLEFISETIQVIQSHEFGEFGADAGGAMRALELFREMANNYELGFGG-CANGENFELLFFG-CANGENFELLFFG-CANGEN
Score = 589.0
Local:
25 MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQC
1 MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQC Score=809
beta-fibrinogen, partial [Equus burchellii quagga]
global:
PRITEIRFILEGICSCPERKOWCY-WCSAVKILIRKYPIECCVQVSFPUV-VKSL-YIINGSLCLIGHKCHKA-VANSSILVIV-NGSFQIP-V-KFREMMYSITTEFKGREINLIMTKYSSSKCALLPUL-LYDG-V7-SL-TSSVW-K-WQLVS-NALE-VKSLUVALTHSKYPHNGSC-W-LSLLIFBNISSDF-SRSEENS
TTTD- PRECENTS- DECOM-

Local:			
205 PRNNISSD			
111			
5 PRKQCSKE			
Score=17			
15 CSCPERKGVW			
11 -11-1			
9 CS-KEDGGGW			
Score=17			
Hemoglobin			
Local:			
153 IPLSWS-LGGHA			
11 VKAAWGKVGAHA			
Score=22			
153 IPLSW-SLGGHA			
11 VKAAWGKVGAHA			
Score=22			
global:			
$\verb TLLVPTDSERTHHGAVSCRQDQRQGRLG*GRRARWRVWCGGPGEDVPVLPHHQDLLPALRPEPRLCPG* $			
		1	
MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKT-YFPHF-DL	SHGSAQVKG	HGKKVADALTNAVAHVDDMPNALSALSDLHA	
${\tt GPRQEGGRRADQRRGARGRHAQRAVRPERPARAQASGGPGQLQAPKPLPAGDPGRPPPRRVHPCGARLPGQVPGFCEHRADLQIPLSWSLGGHASCPLGLPPAPPPLPAPVPPWSLNKV*VGGVPGFCEHRADLQIPLSWSLGGHASCPLGLPPAPPPLPAPVPPWSLNKV*VGGVPGFCEHRADLQIPLSWSLGGHASCPLGLPPAPPPLPAPVPPWSLNKV*VGGVPGFCEHRADLQIPLSWSLGGHASCPLGLPPAPPPLPAPVPPWSLNKV*VGGVPGFCEHRADLQIPLSWSLGGHASCPLGLPPAPPPLPAPVPPWSLNKV*VGGVPGFCEHRADLQIPLSWSLGGHASCPLGLPPAPPPLPAPVPPWSLNKV*VGGVPGFCEHRADLQIPLSWSLGGHASCPLGLPPAPPPPRAPVPPWSLNKV*VGGVPGFCEHRADLQIPLSWSLGGHASCPLGLPPAPPPPRAPVPPWSLNKV*VGG$			
	•		
HKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY	R		

Score=-42