BBM 411

Homework 1

Buğrahan Halıcı

21827483

Q1

a) What is a gene? What is a protein? What is the biological relation between genes and

proteins? What is the difference between a chromosome and genome?

Gene is a small part of a DNA that carries information. Proteins are large molecules composed of amino acids. They regulate all gene activity and provide much of the cellular structure. Genes are responsible for the production of the proteins. The information required to reproduce very complex organisms is stored on a small number of DNA molecules. This molecules is called genome. Chromosomes are the packed structure of a DNA.

b) What is gene expression? What is the name of the process that allows the production of

Manifestation of genes by the operations transcription and genetic translation. multiple versions of a protein from the same gene, and why is this important?

c) Why do we align biomolecular sequences? What do the detected similarities tell us?

To identify regions of similarity that may be a consequence of functional, structural or evolutionary relationships between the sequences. Detected similarities could give us an idea about existence of shared ancestry, descent from a common ancestor, finding motifs, and prediction of functions.

d) What is the use of scoring matrices in the process of alignment? What is the meaning of the

numerical values in a scoring matrix and how are they calculated? Explain and discuss

over the example of BLOSUM62 matrix. What is the meaning of “62” in the name of this

matrix?

Scoring matrices is used for determining the score made by matching two characters in a sequence alignment. Meaning of thenumerical values in a scoring matrix is determined by how similar two residues are. Frequencies are determined by trial and error.

62 stands for: within each block, the amino acid sequences were at least 62% identical when two proteins were aligned.

e) Which one of the following algorithms, BLAST or FASTA, is expected to have a faster run

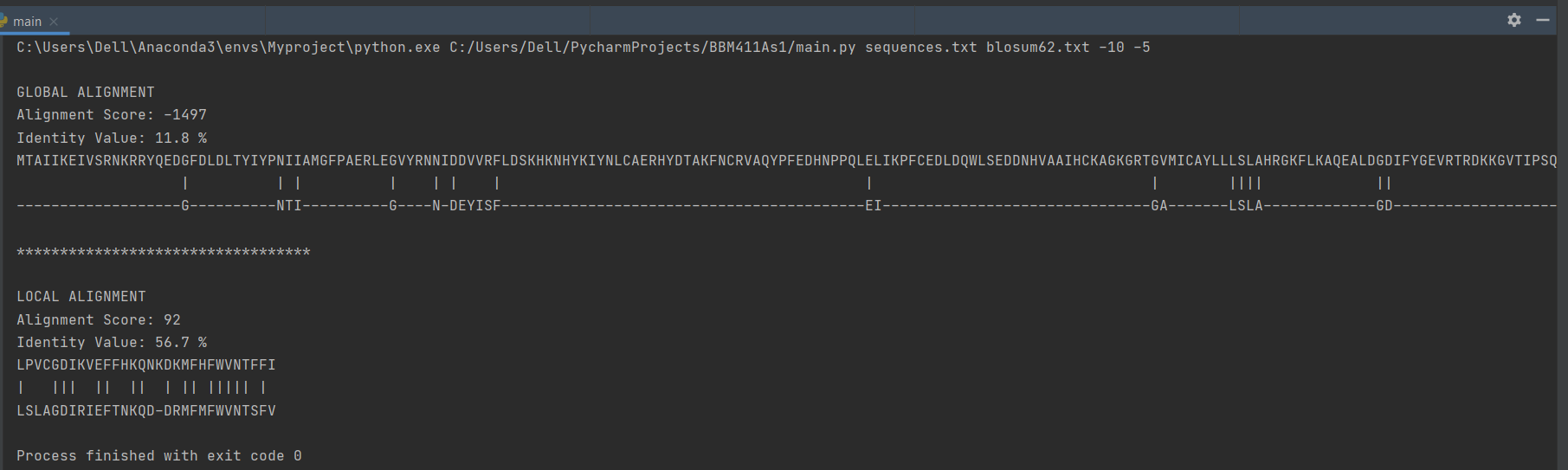
time, given the same search database and query sequence? Which one do you expect to be

more accurate? Explain both over the algorithmic approaches and complexities of these

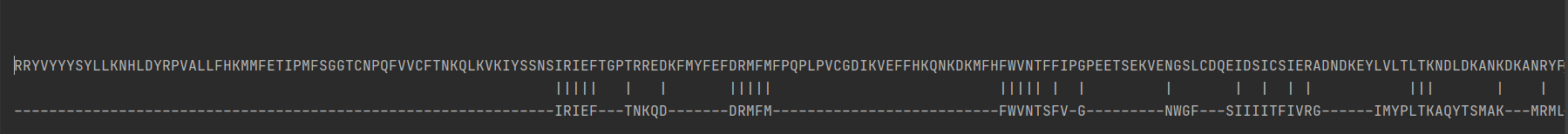
algorithms

BLAST is faster than FASTA by searching for only the more significant patterns in the sequences. FASTA is more accurate than BLAST cause of short DNA and protein sequences in tups.

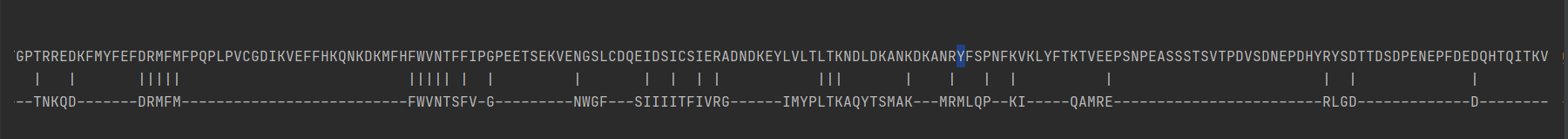
Q2



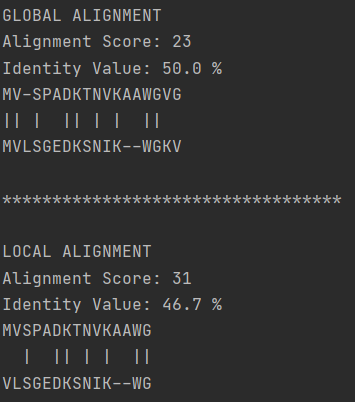
1



1 cont.



end



Sequences :MVSPADKTNVKAAWGVG

MVLSGEDKSNIKWGKV

C) Smith-Waterman algorithm needs to be used. Because if we are looking for the region of functional importance, local alignment finds the local regions with a high level of similarity.

D) If the gap extension penalty decreases, both local and global alignment scores increase. When PAM250 is used as scoring matrix, global alignment score with -1464 increases, global identity value with 10.4 % decreases. Local alignment score with 107 increases, local identity value with 51.5 % decreases.

