

BIO 101 Introductory Biology

Assignment 3

Marks 30

Instructions:

- *Cite the resources you are using properly*
 - *Attach screenshots/pictures in the document and convert it to a single PDF file*
 - *Submit it on LMS*
-
- 1) Download FASTA file for protein ID P05205 from uniprot.
 - a) What is the length of the coding region of the gene that codes for this protein? (2)
 - b) How many exons and introns are present in this protein? What are their sequences? (3)
 - c) How many splice variants are there for this protein? (2)
 - d) What is the function of this protein and what kind of domains are present? (3)
 - e) BLAST this protein in the following organisms. What kind of hits do you obtain? What does that tell you about this protein? (3+3+3) Attach screenshots while describing them as well.
 - i) Homo Sapiens
 - ii) Arabidopsis Thaliana
 - iii) Musca Domestica
 - f) Is the crystal structure of any region of this protein available? What techniques were used to visualize the structure? Attach any available pictures. (4)

 - 2) Find the transition score for the following protein sequences using the BLOSUM62 matrix.
 - a) HENRYCAVILL into HARRYSTYLES (5)
 - b) CHRISPRATT into CHRISEVANS (5)

