**BIO 434 WORKSHEET** 27 POINTS NAME **ANSWER KEY**

Similarity in Developmental Biology – submit the completed worksheet next lab.

**Exercise 1**

Q1. Which of the proteins from the database are the i) most and ii) least similar to our query? (2 pts)

* *The most similar proteins will have the highest scores; the least similar will have the lowest scores*
* *Protein sequence #4, C terminus, is the most similar using* ***MPY*** *as the query word*
* *Protein #2 contains no sequence match to the query word and has the lowest score*

**Exercise 2**

Q2. How similar do you expect the human and mouse myostatin proteins to be? Rationalize your answer. **(3)** (1 pt)

* *Myostatin is a highly conserved protein, as stated in the lab exercise,*
* *one would predict then, that human and mouse myostatin proteins would have very similar sequences (and functions)*

Q3. If we were to compare the *nucleotide* sequences for the gene encoding this protein between humans and the mouse, do you think they would be relatively similar? Explain. **(4)** (2 pts)

* *think of codon degeneracy and the genetic code*
* *there are 64 possible combinations of RNA nucleotides that produce 20 amino acids*
* *knowing this, one would predict that the nucleotides sequences would be similar, as they code for similar amino acids, but not identical, due to the redundancy in amino acid production, with 20 amino acids being produced via 64 possible combinations*

**Exercise 3**

Q4. Were your query and subject sequences identical? Provide evidence for your answer. (2 pts)

* *The query sequence was MSTN H. sapiens, the subject sequence was MSTN Mus musculus*
* *The sequences are not identical, but rather show 86% identity, as indicated on the NCBI website screen shot on the last page*

|  |  |
| --- | --- |
| organism | mRNA base pairs (bp) |
| *H. sapiens* | 2,823  |
| *M. musculus* | 2,705 |

Q5. What percent identity is there between these sequences? Is this at all surprising to you? Explain. **(6)** (2 pts)

* *86% identity at the nucleotide level; 95% identity for amino acids*
* *This makes sense because of the aforementioned codon redundancy*

Q6. Insert a copy of the nucleotide sequence alignment for the mouse myostatin gene from the NCBI website with your submission. (2 pts)

* *There are many different sequences for Mus musculus mRNA in the database*
* *submission of the sequence the student used to answer this question will indicate to you if they interpreted the information on the website correctly*

**General Questions**

Q7. In your own words, define and provide an example of bioinformatics. (2 pts)

* *Computer science and analysis of biological data – DNA, RNA, protein*
* *Example -BLAST -aligning sequences to the NCBI database*

Q8. What is an E value? A sequence alignment returns a high E value - what does this mean? (2 pts)

***Expect value*** *– the probability that the query sequence compared to the database sequence would occur due to random chance. A lower E value indicates a lower chance of random occurrence for the alignment. In essence, the value describes the random back ground noise for a sequence. A high E value indicates that the alignment is likely due to random chance, rather than specific biological features.*

Q9. What is myostatin? (2 pts)

* *Protein signaling molecule (ligand) belonging to the TGF-ß super family. In the lecture we saw that myostatin is a negative regulator of muscle growth – it prevents over-growth of the skeletal musculature.*

Q10. What is the function of myostatin? (1 pt)

* *Myostatin controls muscle cell growth (see Q9 answer)*

Q11. What signaling network does myostatin belong to? ( 1 pt)

* *TGF-ß super family, signaling via intracellular SMAD proteins*

Q12. Summarize the signaling events that occur after a myostatin ligand binds to its cell membrane receptor. ( 2 pts)

* *Myostatin binds activin type II receptor*
* *The type II receptor phosphorylates the activin type I receptor*
* *SMADs 2,3,4 complex and translocate to the nucleus, binding DNA*
* *DNA binding alters gene expression, either gene activation or blockade*

Q13. Knowing that a mutation in the mouse myostatin gene results in muscle hypertrophy, predict what might happen if a similar mutation were to occur in humans. Provide evidence or a rationale for your answer. ( 2 pts)

* *As myostatin is a conserved protein with similar amino acid sequences in mouse and humans, one would predict that same phenotypic outcome in humans as in mice – that is muscle hypertrophy*
* *Mice are used to model human diseases because they have similar molecules and signaling pathways*
* *However, would need further testing*
* *Encourage your students to support their answer with scientific literature, for example:*

*“Myostatin Mutation Associated with Gross Muscle Hypertrophy in a Child”, Markus Schuelke, et al. 2004. New England Journal of Medicine, 350:2682-2688.*

*DOI: 10.1056/NEJMoa040933*

Q14. Using evidence gathered during your Bioinformatics exercise and using myostatin as an example, explain whether or not a mouse might be a good model organism for human development. (2 pts)

* *Mouse and human myostatin share 86% identity at the nucleotide level and 96% at the amino acid/protein level, indicating that they are highly similar. This similarity suggests that the mouse would be a good model for human development.*

Q15. Imagine that a colleague asks you to align a conserved metabolic protein coding gene sequence from a dog to its human homolog. Which sequence type, DNA or protein, would you expect to exhibit the highest percentage of identities? Why? **(7)** (2 pts)

* *There are several different codons that can code for one particular amino acid – codon redundancy. One of the implications of this is that a mutation in the DNA sequence need not alter the protein sequence, if the amino acid inserted is the same. Moreover, mutations in DNA, if they occur in regions that are not required for protein function, could change an amino acid without creating a phenotypic mutation or loss-of-function in the protein product.*

**Example screen shot for Q4.**

****