Name:_____

Activity 4
Protein Data Bank
BIO/CMPSC 300
19 November 2019
Please save this file to act4/ in your repository. Due at the end of class.
GitHub link: https://classroom.github.com/a/DuBJW7yi

1. Go the Protein Data Bank website (http://www.rcsb.org) and search for "Lysozyme" protein. What do you see in your results?

2. Choose one of the proteins and click on its record. Which protein did you pick? There is a four-character alphanumeric code to differentiate one protein from another.

3. Find the Structure Summary tab of this protein. What organism does your protein come from?

4. What is the classification of the protein?

5. What is the FASTA sequence of the protein? You only have to include the first hundred amino acids if this sequence is too long. *Hint: you may have to find this information in another tab.*

6. Back in the Structure Summary tab, what does the literature section appear to say about the protein, in terms of its function? Can you gather any information from the listed articles (and their keywords) about the protein's function here?

7. Go to the "Sequence" tab where the "Sequence Chain View" is shown. What is this information being shown?

8. Are there more chains, sheets, coils or bends (turns) in your sample? Based on what you know about these structures in proteins, is this particular protein going to be flexible or would you conclude that it would be more sturdy? Could your protein be a mixture of both? Why?

9. Attach a screen shot of this protein's Sequence Chain View.

10. Next, go to the "3D View" tab to view the 3D molecular structure of the protein. What do each of the Four Assembly options allow us to see? Make

an effort to describe each view *Hint: you may need to read documentation to inform your discussion*

11. In the display options, what do each of the items of the eight items of the "Colors" menu reveal about the protein? Try each option to learn what they do. Include four screenshots of the most informative graphical outputs (in your opinion), as well as a description of what these particular graphics show. Include why you think these graphical outputs are so informative.

12. Which *Color* option is most informative for determining where the polar and charged regions of the protein are more likely to be found in terms inside or outside of the protein?

13. Discuss the options to show "Water", "Ions" and "Hydrogens" in terms of the protein. What can we learn from these three additions to the protein graphic?

14. Go to the "Sequence Similarity" tab. What is the information provided by this tab? Why is BLAST involved in this tab?

15. How is it possible to determine the function of your protein by consulting types of information in the "Sequence Similarity" tab?

16. At your option: download the PDB file and show a screenshot of the stand-alone Jmol editor displaying the protein. Link to site: http://jmol.sourceforge.net/. Note: **Jmol.jar** was supplied in a previous lesson. This file will need to be run by a Java Virtual Machine. If you would rather not run the JAR file, then please use this time to examine the other options of the webpage's 3D viewer.

17. After taking some time to *play* with your protein in the Jmol viewer or the web page's viewer, discuss why there may be a need to use an external 3D viewer like Jmol to study protein structures when Protein Data Bank's website already provides such a service? Is there any advantage in a stand-alone tool in terms of computational power or another resource? What is different between the tool and the website's tool? If you would rather not run the JAR file, then please speculate on the answers to these questions.

18. Why is it important to be able to visualize proteins? Is having only the amino acid sequence already helpful enough to get an idea about how the protein *functions*?

(Did you remember to put your name at the top of this page?)