	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ PRODUCTON
Jalview 2 **Module Hours:** **2.5**	**Effective Date: 12/1/2023** **PRQs:** **Introduction to Jalview**	**Revision 1.0** **Author:** **D. Roberts** **Checked by** **Editor: M. Stowell**

Background:

Sequence alignment in a nutshell is simply the process of comparing biological sequences. These "similarities" being detected refer to the number of identical residues/nucleotides across sequences.

Researchers have used sequence editing for decades to determine the evolutionary heritage of organisms. Over the course of evolution, much of the code for life is maintained. The DNA polymerase of a mouse has a near identical sequence to the DNA polymerase of a human. A DNA polymerase in an E. Coli will have a similar sequence to a human but not as similar as a mouse. This data shows us that a mouse is more closely related to a human than an E. Coli bacterium, but it also points to a common "code of life" inferring that all organisms are related. Sequence alignment helps us determine how closely related we are.

In addition to evolutionary study, sequence alignment helps us compare structures to determine functions. An **alignment score** measures the number of residues that are identical in the sequences being compared. If the alignment score is high, the sequence is closer to being identical. A more highly related sequence often points to a similar structure, and consequently, a similar function. If we were studying an abnormal protein found in stomach cells, but discovered its sequence alignment was very similar to a proteasome, we could infer that the protein probably has a function that is similar to a proteosome. While this strategy is often effective, comparing actual structure rather than just alignment score is also necessary to form these inferences. Sometimes, 2 proteins could have a completely different sequence but a very similar structure, and 2 proteins could have a very similar sequence but have a very different structure, depending on the properties of the residues. However, alignment score is a piece of the puzzle that can be used to draw conclusions.

	SKILLS CENTER **STANDARD OPERATING PROCEDURE**	**A BIOFIZZ** **PRODUCTON**
Jalview 2 **Module Hours:** **2.5**	**Effective Date: 12/1/2023** **PRQs:** **Introduction to Jalview**	**Revision 1.0** **Author:** **D. Roberts** **Checked by** **Editor: M. Stowell**

While incredibly applicable, protein sequencing is difficult to do without a program. Manually comparing protein sequences is tedious and costly. However, with a computer program to line up the data for us, it becomes easier and more manageable to run sequence alignments.

Jalview is a free, cross-platform program used for multiple sequence alignment editing, visualization and analysis. First developed in 1995 by the University of Oxford, Jalview has been run by the University of Dundee since 2001. In addition to its sequence alignment capabilities, Jalview has built in DNA, RNA and protein sequence and structure visualization and analysis capabilities. It provides a linked view of aligned DNA and Protein products as well as their structures and family trees.

1. Purpose

The purpose of this procedure is to recognize the multitude of applicable uses of Jalview sequence analysis and to familiarize oneself with its most common applications.
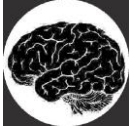
2. Scope
- This procedure applies to qualified skills center users.

3. Responsibility
- It is the responsibility of the user to understand and perform the procedure described in this document.
- It is the responsibility of the user performing the procedure to fully document any deviations from the written procedure (in a computer lab, document your findings using the database.)
- It is the responsibility of the user to become trained in the use of this application.

4. Definitions
- RCSB Protein Data Bank – A database that houses information about 3D macromolecule structures, such as proteins and nucleic acids.
- Residue – an amino acid within a polypeptide sequence (protein)
- Sequence – chain of amino acids

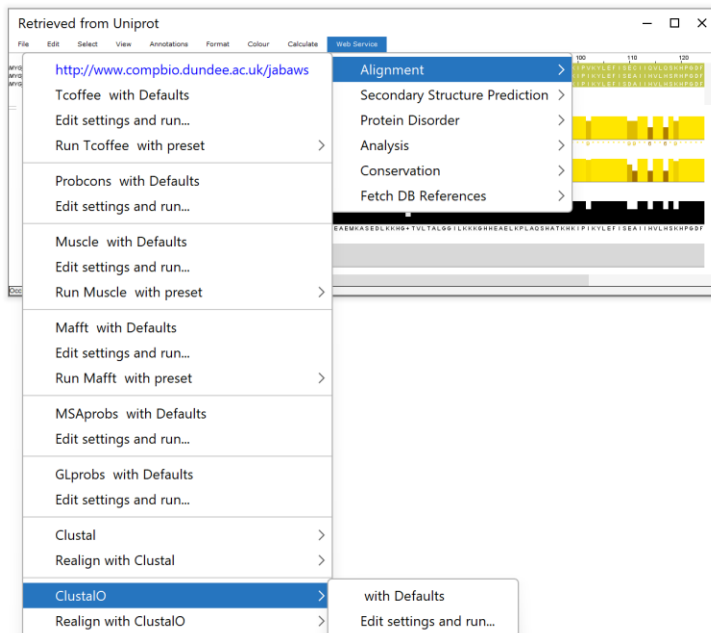|	**SKILLS CENTER** **STANDARD OPERATING PROCEDURE**	**A BIOFIZZ** **PRODUCTON**
Jalview 2 **Module Hours:** **2.5**	**Effective Date: 12/1/2023** **PRQs:** **Introduction to Jalview**	**Revision 1.0** **Author:** **D. Roberts** **Checked by** **Editor: M. Stowell**

- JalView – an application that allows for detailed protein, DNA, and RNA sequencing, structure visualization, and tree creation.
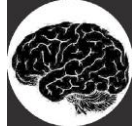
5. Materials/Equipment
- Jalview Application
- RCSB PDB

6. Procedures
- **6.1 Aligning your protein sequences:**
  - ○ Open your Jalview application with your desired proteins from the Introduction to Jalview module.
  - ○ Before we can analyze the sequences, we must align them. Select "Web Service" in the alignment window and then "Alignment" -> "ClustalO" (the method we will be using to align our sequences for this module) -> "With Defaults."
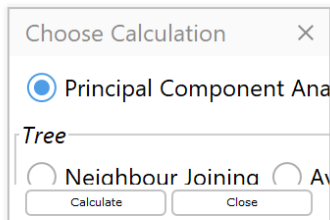
	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ PRODUCTON
Jalview 2 **Module Hours:** **2.5**	**Effective Date: 12/1/2023** **PRQs:** **Introduction to Jalview**	**Revision 1.0** **Author:** **D. Roberts** **Checked by** **Editor: M. Stowell**

- **6.2 Principal Component Analysis:**
  - o Now that your sequences are aligned, you can analyze them: Select "Calculate" in the alignment window tab and select "Calculate Tree or PCA."
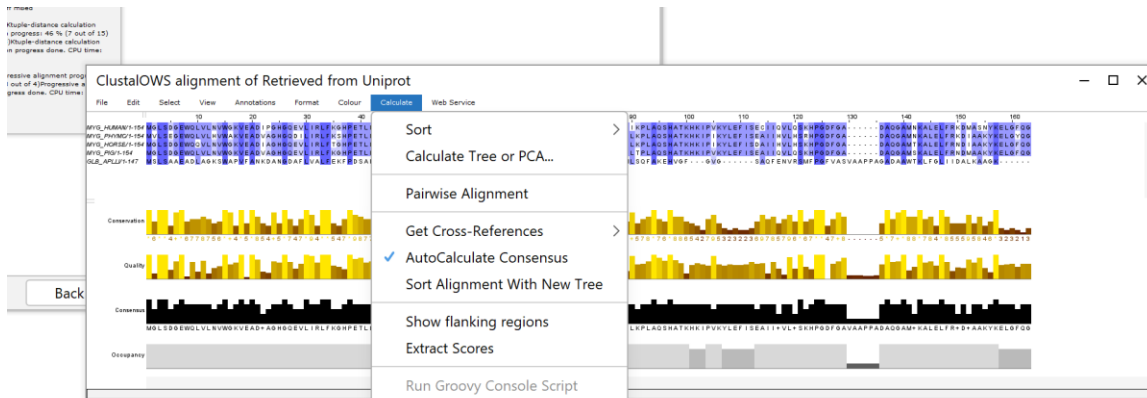

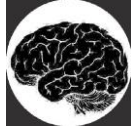 - o In the menu that pops up, select "Principal component analysis."

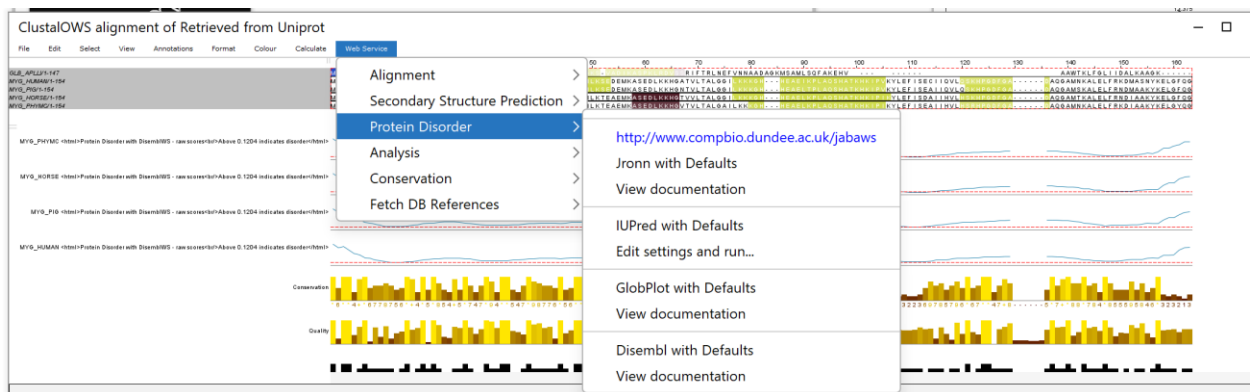- o A 3D graph will appear with a visual representation of how closely each protein is related. Each "white box" represents a protein and you can identify which protein it is by hovering your mouse over it. To label every protein, select "view" -> "show labels."
- o You can change the background color of the chart by selecting "view" -> Change background color.
- **6.3 Tree Display:**
  - o To make a family tree of your sequences, in your alignment sequence tab, select "Calculate" -> "Calculate Tree or PCA" and select the "Neighbor Joining" option in the pop-up window.
  - o To change the distance display settings, select the "view" tab on the tree window, and select "show distances."
- **6.4 Structure Display:**
  - o You can view a 3D Structure of your sequence by selecting the sequence in the alignment window, right clicking, and selecting "3D Structure Data."
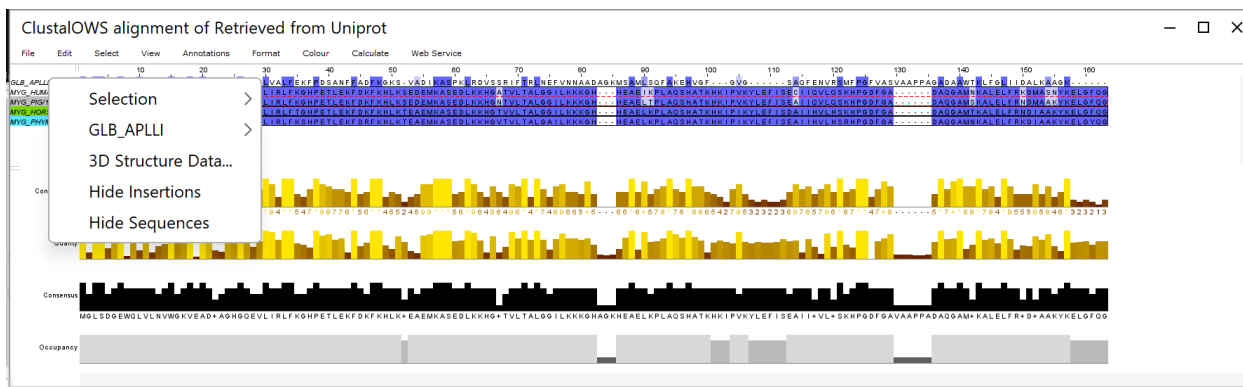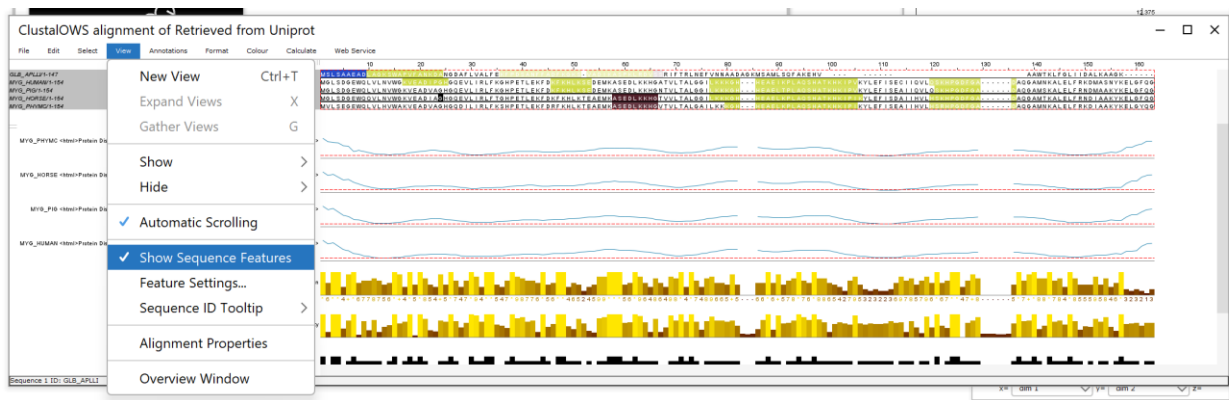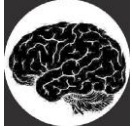

- o Clicking a specific region of your resulting structure will highlight the specific residue in your sequence.
- **6.5 Protein Disorder Predictions:**
  - o First, turn off all color indicators for your sequences by selecting "colour" -> "none." Next, and select "Web Service" followed by "Protein disorder" -> "Disembl with defaults."

	SKILLS CENTER **STANDARD OPERATING PROCEDURE**	**A BIOFIZZ** **PRODUCTON**
Jalview 2 **Module Hours:** **2.5**	**Effective Date: 12/1/2023** **PRQs:** **Introduction to Jalview**	**Revision 1.0** **Author:** **D. Roberts** **Checked by** **Editor: M. Stowell**

- o A new section should appear in your alignment window. This is a graph of likelihood of disorder. A higher score (or a higher blue line) indicates a greater likelihood that this section of the sequence will retain a protein disorder.
- o There are multiple factors that go into the likelihood of a section of a sequence retaining a disorder. Select "View" -> "Show Sequence Features" followed by "View" -> "Sequence Feature Settings" to view the other factors.

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ PRODUCTON
Jalview 2 **Module Hours:** **2.5**	**Effective Date: 12/1/2023** **PRQs:** **Introduction to Jalview**	**Revision 1.0** **Author:** **D. Roberts** **Checked by** **Editor: M. Stowell**

- o A new window should pop up containing the three factors that go into a disorder calculation. The REMARK465 annotation marks regions where the coordinates in the Xray structures are missing, indicating a higher likelihood of a disorder. In addition to the REMARK465 Score, coils and "hotloops" also impact the likelihood of a disorder. Coils are areas of the sequence where the structure is supercoiled, and "hotloops" are a subcategory of a coil but a particularly mobile coil. Both coils and hotloops are also common indicators of a protein disorder.
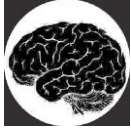
7. Troubleshooting
- If the application will not run on your PC because it is labelled as "not safe," click "more information" and then at the bottom of the pop up, click "run anyway"
- You might need to update the app to the existing installation: just follow the prompts.
- The Jalview website has some good training videos on the home page that might be helpful if you are confused with any of the tasks.

8. References

"Jalview Home Page." *Jalview*, www.jalview.org/. Accessed 13 Oct. 2023.

9. Module Mastery Task (MMT)
- 9.1 Align your sequences and submit a resulting image. What changed from your original alignment? Why is it important to align sequences?
- 9.2 Performa a Principal Component Analysis test on your sequences. What do the closer coordinates tell you about their sequence similarities?
- 9.3 Label the proteins on the analysis and change the color of the background. Submit the resulting image.
- 9.4 Generate a tree of your sequences and submit a resulting image with the distance labels OFF. Did anything surprise you about how these proteins are related? What could you do to confirm the accuracy of this tree?
- 9.5 Generate a 3D structure for each sequence and line them up in the same orientation side by side (should be 5 separate windows). What do you notice about each structure? Are the structures similar? What do the structures of the proteins tell you about their function?
- 9.6 Run a protein disorder analysis. Label areas of coils, hotspots, and REMARK 465. Submit an image of your sequence with locations danger highlighted. Write a paragraph summarizing the effects of a sequence disorder of a protein. Why is it important to locate disorder-prone areas of a sequence? Why do you think a coil, or a hotspot would increase the likelihood of a disorder?
- 9.7 Write a concise conclusion that encompasses the following key elements:

	SKILLS CENTER **STANDARD OPERATING PROCEDURE**	**A BIOFIZZ** **PRODUCTON**
Jalview 2 **Module Hours:** **2.5**	**Effective Date: 12/1/2023** **PRQs:** **Introduction to Jalview**	**Revision 1.0** **Author:** **D. Roberts** **Checked by** **Editor: M. Stowell**

- o Restate the purpose of the Module Mastery Task (ie. the objectives you were expected to achieve).
- o Describe the most significant lessons or insights you gained from completing the tasks.
- o Explain how you intend to apply the knowledge and skills acquired during this module in your academic or professional pursuits.

10. Feedback
- This is your space to help make this module better! Which aspects of the module did you like? Which aspects can be improved upon? Do you have any remaining areas of confusion? Did the number of hours assigned to this SOP accurately reflect the time spent on this skill?