BACKGROUND

Proteins are the most complex biological macromolecule, comprised of a variety of amino acids with unique structures and properties which give regions of the protein its characteristics, such as catalytic ability and binding sites, and ultimately determine the three-dimensional shape of the protein. While a protein's shape influences its function and behavior, the key cellular interactions it has and processes which it is involved in is dependent on specific regions of the protein. A protein may have several unique regions which serve different cellular functions, allowing for interconnection of pathways and signal integration. The nature of proteins leads to complex possibilities. By studying the functions belonging to a certain protein and what properties of the protein contribute to that function, molecular biologists can not only gain greater understanding of ways to moderate and change the protein's functions, but also compile data which can be used to assign functions to other proteins based on similar structures or functional regions.

Many databases are dedicated to the collection of large amounts of data on proteins. Specifically, the protein interaction databases STRING and STITCH compile information on interprotein interactions and protein-chemical interactions and use known information to predict interactions. The STRING database includes information on interprotein interactions. STRING stands for Search Tool for the Retrieval of Interacting Genes/Proteins. The STITCH database includes information on protein interaction with other small molecules. The STITCH acronym is short for Search Tool for Interactions of Chemicals. The Swiss Institute of Bioinformatics (SIB), European Molecular Biology Laboratory (EMBL), and the Novo Nordisk Foundation Center for Protein Research (NNF-CPR) come together to form both the STRING and STITCH consortiums (STRING, 2020). These two databases harbor information on protein interactions from a multitude of different sources.

As do other protein databases, STRING and STITCH compile interaction information from several different literature sources to create an all-encompassing information bank. Information is compiled from cited literature, databases of protein interaction experiments and annotated protein pathways, computational interaction predictions, and the transfer of interaction evidence from one organism to another (Szklarczyk, et al., 2020). Some of the databases information is extracted from include the Protein Data Bank (PDB), PubChem, and the Reactome (Kuhn, et al., 2007). This module will explore the use of the STRING and STITCH programs to assess protein interaction information.

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ PRODUCTON
Using STRING and STITCH **Module Hours 2**	**Effective Date: 03/22/21** **Checked: A. Siclair**	**Revision # 1.0** **M. Guzie**

1. PURPOSE

The purpose of this procedure is to learn how to use the STRING and STITCH databases and collect interaction data on proteins.

2. SCOPE

This procedure applies to qualified skills center users.

3. RESPONSIBILITY
3.1.   It is the responsibility of the user to understand and perform the procedure described in this document.
3.2.   It is the responsibility of the user performing the procedure to fully document any deviations from the written procedure.
3.3.   It is the responsibility of the user to become trained on the procedure.

4. DEFINITIONS
4.1. Query Molecule: The molecule which is being searched in the database.
4.2. Consortium: A grouping together or association.
4.3. Node: A spherical shape used to represent a protein in the STRING and STITCH network views.
4.4. Edge: A line used to connect protein nodes which represent specific interactions in the STRING and STITCH network views.

5. MATERIALS/EQUIPMENT
5.1.   The STRING database: https://string-db.org/.
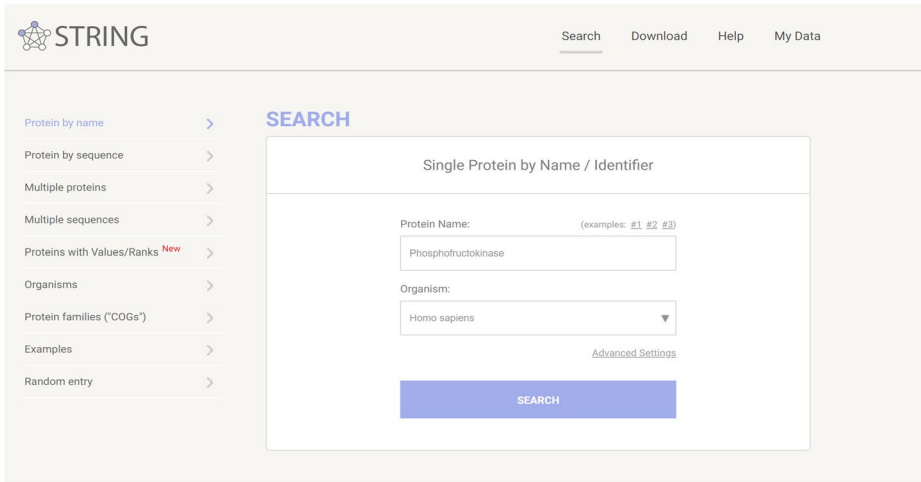5.2.   The STITCH database: http://stitch.embl.de/.

6. PROCEDURE

This procedure will describe how to access annotation data on both protein-to-protein interactions and protein-to-molecule interactions by exploring both the STRING and STITCH databases.

6.1. Using the STRING database
6.1.1. Go to the STRING database: https://string-db.org/.
6.1.2. Type the protein of interest's name into the search bar.

	SKILLS CENTER **STANDARD OPERATING PROCEDURE**	**A BIOFIZZ** **PRODUCTON**
Using STRING and STITCH **Module Hours 2**	**Effective Date: 03/22/21** **Checked: A. Siclair**	**Revision # 1.0** **M. Guzie**

6.1.3. Select an organism to search from the drop-down bar.

Figure 1: The STRING homepage with input ready to run a protein-to-protein interaction search for the protein Phosphofructokinase in homo sapiens.

6.1.4. Once a protein name and an organism have been selected, select search. (**Figure 1**)

	SKILLS CENTER **STANDARD OPERATING PROCEDURE**	**A BIOFIZZ** **PRODUCTON**
Using STRING and STITCH **Module Hours 2**	**Effective Date: 03/22/21** **Checked: A. Siclair**	**Revision # 1.0** **M. Guzie**

6.1.5. The search may result in several matches in the database. Make sure that the desired protein is selected, and then press continue. (**Figure 2**)

Figure 2: The list of results which matched the original protein/organism search. Here, the user can select the matching result which they would like to view (Option 1 selected.)

![CU Buffalo logo]	**SKILLS CENTER** **STANDARD OPERATING PROCEDURE**	**A BIOFIZZ** ![brain logo] **PRODUCTON**
Using STRING and STITCH **Module Hours 2**	**Effective Date: 03/22/21** **Checked: A. Siclair**	**Revision # 1.0** **M. Guzie**

6.1.6. The resulting page will yield a web of interconnected proteins which have interactions with the query protein. Each protein is represented by a colored node (sphere) while each interaction is represented by a colored edge (string). Select protein nodes to pull up information on that individual protein, including functional information, identifier codes, and organism. More actions can also be selected from here, such as showing the sequence, finding homologs, and other related pathways and functions. A structure is also shown, gathered from either a structural database such as the protein data bank (PDB) or a homology modeling algorithm. Selecting on the structure will bring the user to the database from which it is compiled. Also select interaction edges to find information on the interaction between the two connected proteins along with evidence supporting the interaction. (**Figure 3**)

Figure 3

Figure 3: The STRING network results for the protein Phosphofructokinase (represented by PFKL, red node pulled to the left.) By selecting the protein nodes, information gathered on the individual protein will be displayed. By selecting the interaction edges, information on the interactions between the two connected proteins will be displayed.
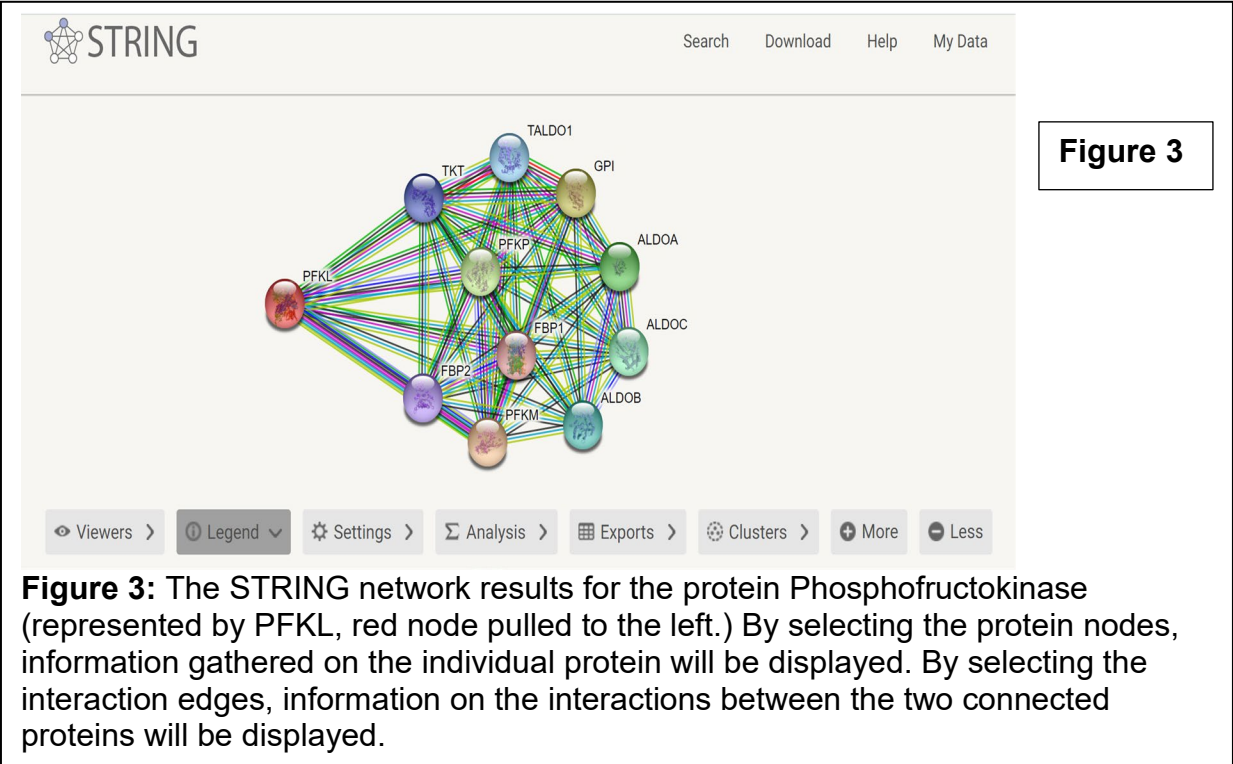
6.1.7. By selecting Legend and scrolling down, a figure legend will be provided indicating the importance of each color in the STRING network for both nodes and edges. A
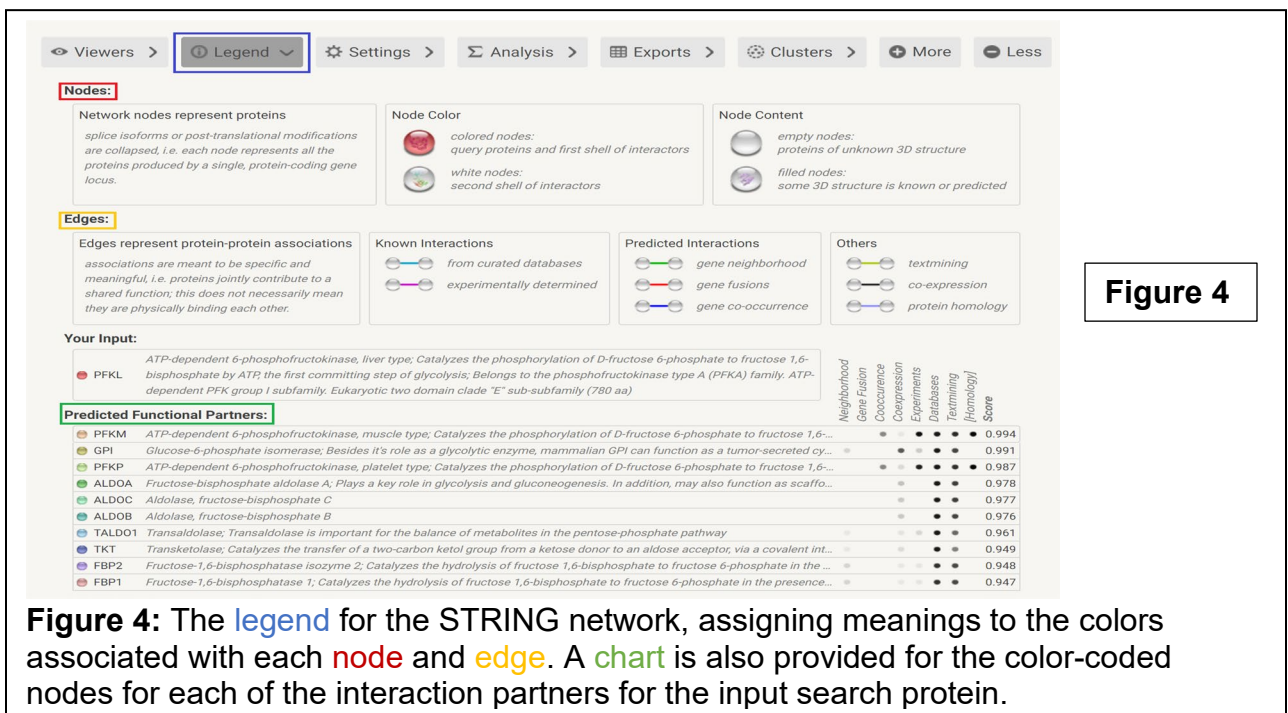
chart of the color-coded nodes for the specific search results will also be provided along with their general functional role. (**Figure 4**)

Figure 4: The legend for the STRING network, assigning meanings to the colors associated with each node and edge. A chart is also provided for the color-coded nodes for each of the interaction partners for the input search protein.

6.1.8. There are several other tabs besides the legend tab that can be used for filtering and interpreting results.

-**Settings** allows the user to adjust what exactly the results are displaying and how they are being displayed. The user can adjust which sources are being used to draw interaction information from and how the figure is displaying the data.

-**Analysis** gives a statistical overview of the results, including general information on the query protein's interactions with other proteins and also links other databases which give information on metabolic pathways and protein domains associated with the search.

-**Exports** allows the user to download the STRING network in several different formats.

|	**SKILLS CENTER** **STANDARD OPERATING PROCEDURE**	**A BIOFIZZ** **PRODUCTON**
Using STRING and STITCH **Module Hours 2**	**Effective Date: 03/22/21** **Checked: A. Siclair**	**Revision # 1.0** **M. Guzie**

-**Clusters** allows the user to select one of two different possible means of clustering to group more similar proteins together.

-**More** will add protein nodes to the STRING network, creating a denser web of more interactions.

-**Less** will remove protein nodes from the STRING network, creating a simpler web with less interactions.

6.1.9. After using the legend and other tabs to gain an understanding of the STRING network, move to the Viewers tab to view other options in which the protein-to-protein interaction network can be viewed. There are 8 different options to view the interaction information in different formats, including the standard network summary view (explored above), an experimental view imported from primary resources, automated text mining for proteins often mentioned in published literature together, several different ways to group genes, and co-expression. Note that for each of these different viewpoints, there is also a legend, settings, analysis, and table tab as for the standard network viewpoint. (**Figure 5**)
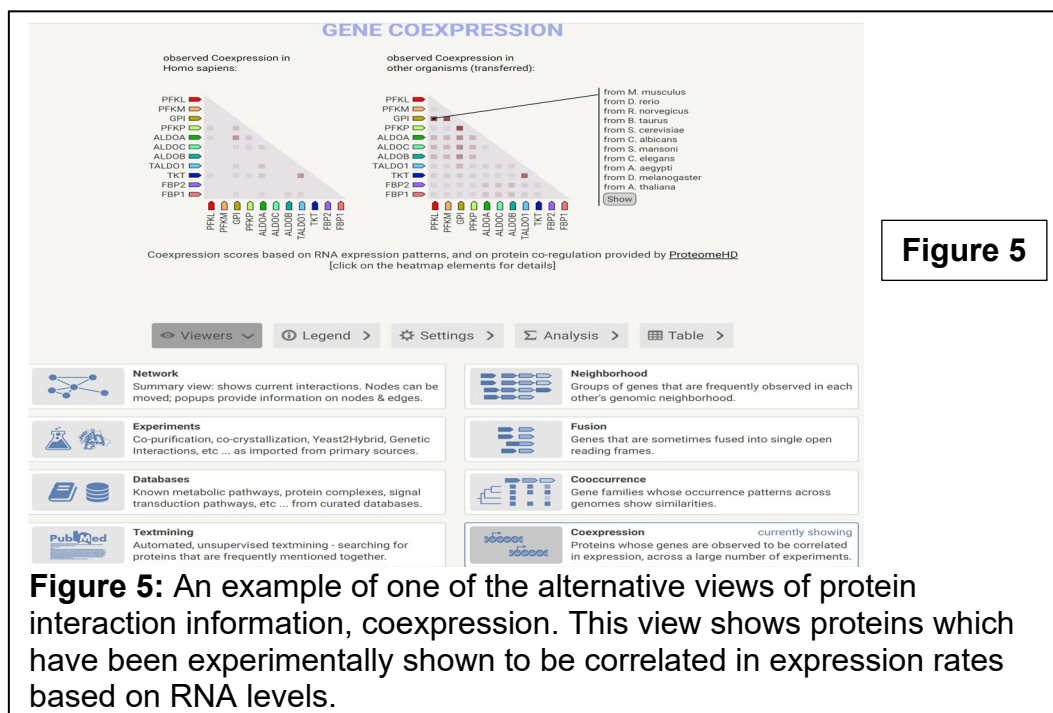
Figure 5

Figure 5: An example of one of the alternative views of protein interaction information, coexpression. This view shows proteins which have been experimentally shown to be correlated in expression rates based on RNA levels.

6.2. Using the STITCH Database

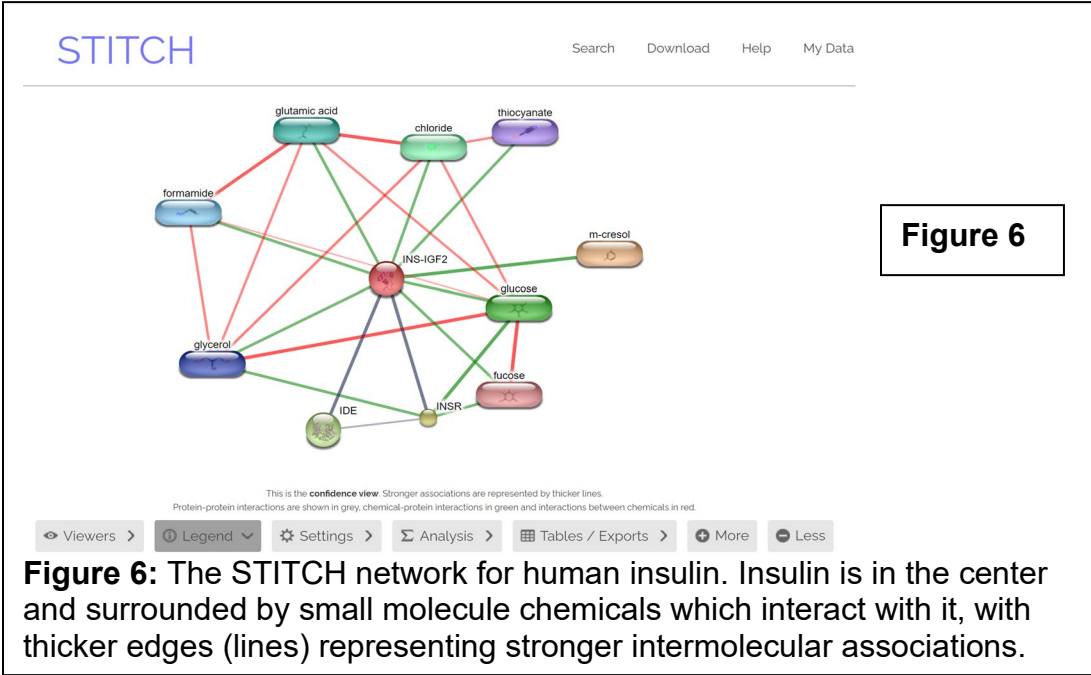| ![CU Buffalo logo]	**SKILLS CENTER** **STANDARD OPERATING PROCEDURE**	**A BIOFIZZ** ![brain logo] **PRODUCTON**
Using STRING and STITCH **Module Hours 2**	**Effective Date: 03/22/21** **Checked: A. Siclair**	**Revision # 1.0** **M. Guzie**

6.2.1. Go to the STITCH database: http://stitch.embl.de/.

6.2.2. Protein-to-chemical interactions can be searched through STITCH by searching for a protein OR a chemical. By searching a protein, the results will yield mostly chemicals which have interactions with this protein but may also show some protein interactions. By searching a small molecule chemical, the results will yield mostly proteins which the chemical has interactions with. Run a search for a protein first to compare how the STITCH network for a protein looks vs. the STRING network for a protein.

6.2.3. Select an organism to search from the drop-down bar.

6.2.4. Once a protein name and an organism have been selected, select search.

6.2.5. The search may result in several matches in the database. Make sure that the desired protein is selected, and press search.

6.2.6. The resulting STITCH network will have the same form as the STRING network, with colored nodes for proteins/chemicals and colored edges attaching them to show their interactions. The network will also have the same general functions as the STRING network: select nodes and edges to view more information about the molecule or interaction of interest. (**Figure 6**)
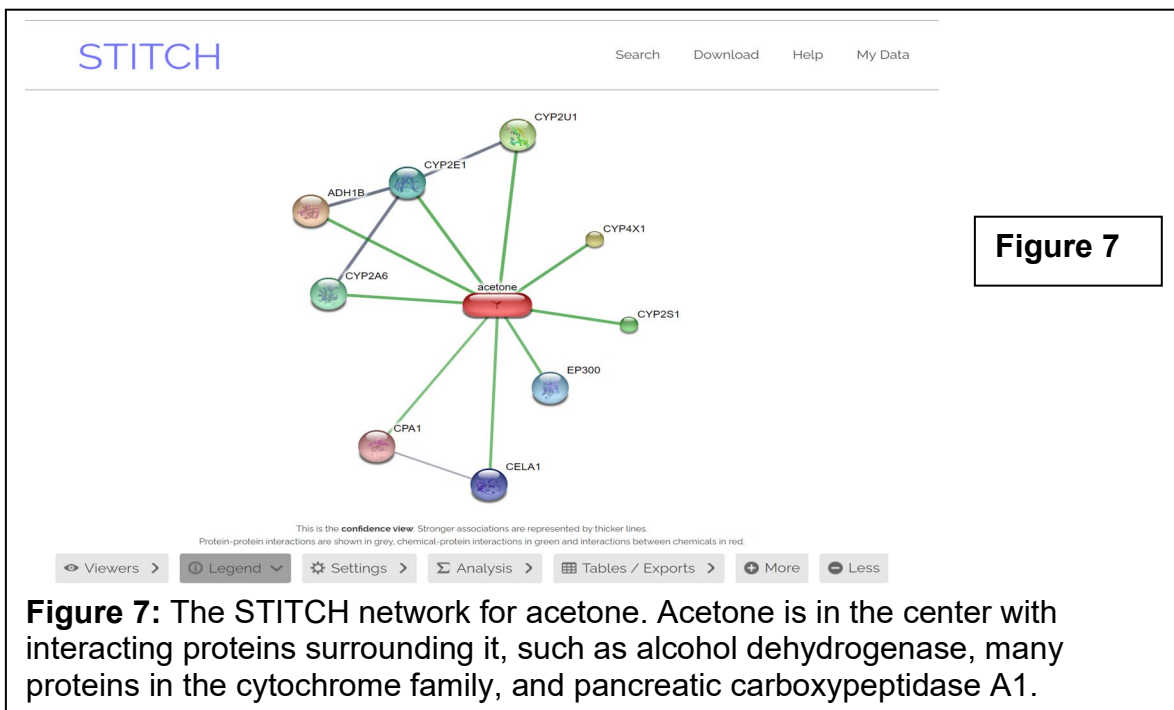

Figure 6

Figure 6: The STITCH network for human insulin. Insulin is in the center and surrounded by small molecule chemicals which interact with it, with thicker edges (lines) representing stronger intermolecular associations.

CU Logo	**SKILLS CENTER** **STANDARD OPERATING PROCEDURE**	**A BIOFIZZ** Brain Logo **PRODUCTON**
Using STRING and STITCH **Module Hours 2**	**Effective Date: 03/22/21** **Checked: A. Siclair**	**Revision # 1.0** **M. Guzie**

6.2.7. Select the viewers tab and go through the other options in which the intermolecular interactions can be displayed. Note that most of the viewpoints and exploratory functions are the same as the STRING database.

6.2.8. Return to the STITCH homepage and run a search on a chemical rather than a protein.

6.2.9. The resulting STITCH network will have the chemical in the center surrounded by proteins which it has interactions with. View some of the proteins

	Figure 7

Figure 7: The STITCH network for acetone. Acetone is in the center with interacting proteins surrounding it, such as alcohol dehydrogenase, many proteins in the cytochrome family, and pancreatic carboxypeptidase A1.

which interact with the query chemical and view the interaction data in different viewpoints. (**Figure 7**)

	SKILLS CENTER **STANDARD OPERATING PROCEDURE**	**A BIOFIZZ** **PRODUCTON**
Using STRING and STITCH **Module Hours 2**	**Effective Date: 03/22/21** **Checked: A. Siclair**	**Revision # 1.0** **M. Guzie**

7. TROUBLE SHOOTING

8. REFERENCES

Kuhn, M., Von Mering, C., Campillos, M., Jensen, L., & Bork, P. (2007, December 15). *Stitch: Interaction networks of chemicals and proteins.* National Center for Biotechnology. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2238848/.

Szklarczyk D, Santos A, von Mering C, Jensen LJ, Bork P, Kuhn M. STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. Nucleic Acids Res. 2016 Jan 4;44(D1):D380-4.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, von Mering C. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019 Jan; 47:D607-613.

Szklarczyk, D., Gable, A., Nastou, K., Lyon, D., Kirsch, R., Pyysalo, S., . . . Von Mering, C. (2020, November 25). *The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets.* Nucleic Acids Research. Retrieved from https://academic.oup.com/nar/article/49/D1/D605/6006194

	SKILLS CENTER **STANDARD OPERATING PROCEDURE**	**A BIOFIZZ** **PRODUCTON**
Using STRING and STITCH **Module Hours 2**	**Effective Date: 03/22/21** **Checked: A. Siclair**	**Revision # 1.0** **M. Guzie**

9. **MODULE MASTERY TASK**

   9.1. Go to the STRING database and run a search for a protein of interest. Make a note of at least 5 proteins your protein of interest interacts with and what these interactions entail. Take a screenshot of the STRING network and save the photo for submission. Do this for 3 other formats to view the protein-protein interactions in as well, including a figure legend for each viewpoint describing what the viewpoint is showing about this protein's interactions with other proteins.

   9.2. Go to the STITCH database and run a search for a small chemical of interest. Make a note of at least 5 proteins your chemical of interest interacts with and what these interactions entail. Take a screenshot of the STITCH network and save the photo for submission. Do this for 3 other formats to view the chemical-protein interactions in as well, including a figure legend for each viewpoint describing what the viewpoint is showing about this chemical's molecular interactions.

   9.3. As a scientist working on cancer research, you aim to identify proteins which may be involved in the onset and perpetuation of cancer. One focus in your research is on the activity of RAS, a g-protein whose up-regulation can initiate cellular divisions. Loss of regulation of the RAS pathway can lead to uncontrolled cell division. Use the STRING database to search for other proteins which have interactions with RAS and may play a role in the pathway which leads to mitosis. Identify 2 potential candidates and record them.

   9.4. Use the STITCH database to identify molecules which have interactions with the 2 proteins identified from the STRING search. For each protein, make record of 2 chemicals which have interactions with your protein, and how the molecular interaction could be used to your advantage in the study of cancer (could it inhibit a protein involved in the mitosis pathway? Activate a protein involved in DNA repair? Etc.)