Modeling Protein-Protein Interactions using Co-fractionation Mass Spectrometry (CF/MS) Data

My My Tran[1,2], Ahmed Youssef[3,4], Andrew Emili[4]

[1]Smith College, Department of Biological Sciences, [2]Boston University Bioinformatics BRITE REU Program, Summer 2022,
[3]Graduate Program in Bioinformatics, Boston University, [4]Boston Center for Network Systems Biology, Boston University School of Medicine

Introduction

Most proteins carry out their essential cellular functions through sets of physical interactions collectively known as protein-protein interaction networks (PPINs). PPINs are dynamic systems and can be influenced by external conditions that lead to disease-causing changes. Understanding PPINs is a crucial step toward the development of protein biomarkers for drug discovery and therapeutics. Co-fractionation mass spectrometry (CF/MS) is a powerful method for high-throughput identification of protein interactions but stands to benefit from dedicated data analysis tools. Thus, the goal of this project was to develop an interactive web application for visualizing CF/MS data using the *R Shiny* package. This would streamline the process of analysis of CF/MS data for researchers. An *E. coli* dataset that was generated from the Emili Lab was used as a case study for testing purposes.

Background and Motivation

CF/MS identifies the proteins that are present in a sample and their abundance across fractions, where abundances are indicated as intensity. Protein complexes can be predicted using correlations between fractionation profiles.

The application provides visualizations and analyses for researchers to make hypotheses and ask questions based on their CF/MS data. Some examples of these questions include:

- How might protein interactions differ across tissues (heart, lungs, brain, etc.)?
- What are the proteins that may be conserved within a complex? (Evolutionary biology)
- Which interactions are disrupted due to diseases (e.g. cancer)?

Methods

1. *E. coli* datasets were used as input into the *R Shiny* application. The CF/MS data is in the format *MaxQuant.*

2. Replicate files are accepted. The backend stores file inputs as a list of datapaths that can be iterated using the *lapply* function.

3. A helper script pre-processes the CF/MS data using the *tidyverse* package into "tidy" format.

4. Different tabs to visualize fractionation profiles of one or more proteins from the data using the *ggplot2* package.

5. The data is transformed into wide format and heatmap visualizations are created using the *pheatmap* package.

Results and Demonstration

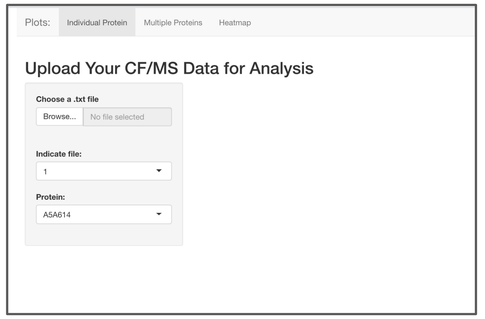

Figure 1. The user interface (UI) for researchers to upload their CF/MS Data. The ShinyWidget browse button searches for local files.

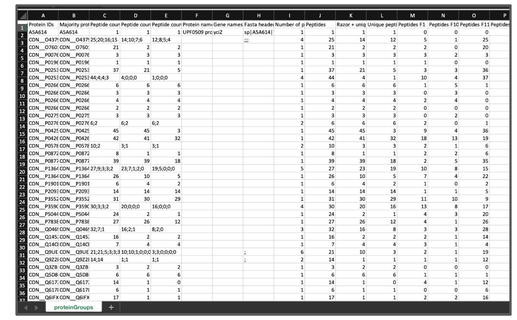

Figure 2. This is an example of CF/MS data that is used as input. This is from an experiment studying the *E.coli* interactome.

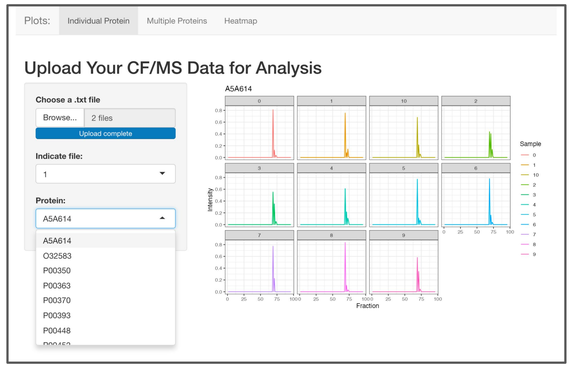

Figure 3. This tab allows the user to subset by replicates and proteins to view the fractionation profiles of individual proteins.

Figure 4. This next tab allows the user to subset by replicates and select multiple proteins. The fractionation profiles of each protein are overlaid.

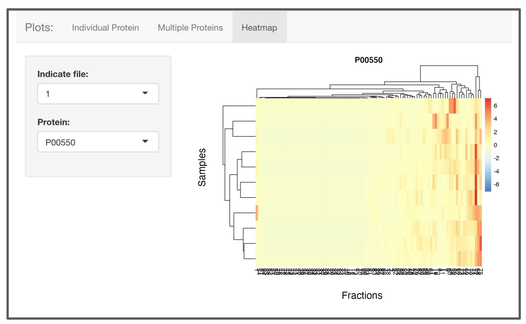

Figure 5. In the heatmap tab, the user can select the replicate file and the individual protein they would like to view across samples and fractions.

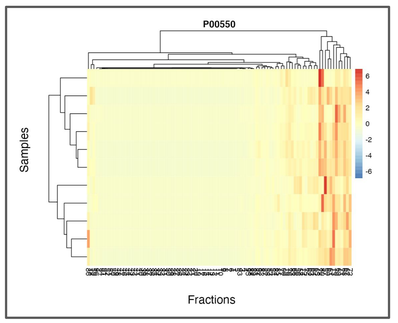

Figure 7. This is a heatmap of the same protein (P00363) from fig.6, however this is the from the second replicate. Note the similarities and differences between the experiment replicates.

Future Directions

- Build PPI networks by calculating distance between proteins and scoring protein similarity.
- Use peptide-level data.
- Reduce run-time by parallel processing.

References and Acknowledgements

- Dupree EJ, *et al.* A *Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of this Field.* doi: 10.3390/proteomes8030014.
- McWhite CD, *et al. Co-fractionation/mass spectrometry to identify protein complexes.* https://doi.org/10.1016/j.xpro.2021.100370.
- Koh GC, *et al. Analyzing protein-protein interaction networks.* doi: 10.1021/pr201211w.
- RStudio. *Hex-stickers.* https://github.com/rstudio/hex-stickers/.

This work was funded, in part, by NSF grant DBI-1949968, awarded to the Boston University Bioinformatics BRITE REU program.