

Abstract

This is a collaborative project with Dr. Szymanski's group in Agronomy to predict protein complex assemblies and interactions. Proteins in the leaf cytosol of arabidopsis were fractionated using Size Exclusion Chromatography (SEC) and mixed-bed Ion Exchange Chromatography (IEX). Protein mass spectrometry data were obtained for the two platforms of separation and two replicates of each. We combine the four data sets and conduct a series of advanced clustering analyses, including 1) data filtering, 2) a two-round hierarchical clustering to integrate multiple data types, 3) validation of clustering based on known protein complexes, 4) mining dendrogram trees for subsets of proteins with high confidence of complex prediction. Our method is developed for integrative analysis of different data types and it eliminates the difficulty of choosing an appropriate cluster number in clustering analysis. It provides a computational tool to globally analyze the oligomerization state of a system of protein complexes.