Computational Interaction Proteomics: from Proteome to Complexome

J. Kutzera

# Summary

Proteins are biomolecules and play important roles in every living cell. They perform or facilitate nearly all cellular functions, including metabolism, replication of the genetic material, communication within the cell, and transport of intracellular molecules. Proteins perform their functions in close contact with each other, and such assemblies of interacting proteins are called protein complexes. Those complexes are the cellular machines that carry out the functions.

Protein interactions can be categorized into different levels. At the lowest interaction level, units are built of firmly bound proteins that exclusively occur together. These units are called "complex–cores" and represent the "LEGO–bricks" of the protein–complex world. All complexes are built from one or more of such complex–cores, and as with the famous Danish toy bricks, each complex–core can be part of different larger complexes. These larger complexes represent protein–protein interactions at higher levels. The larger complexes also interact with each other and temporarily build even larger assemblies, which represent the highest interaction level.

Detecting protein complexes and their interactions is crucial for a better understanding of the cell. A laboratory method called "immunoprecipitation followed by liquid chromatography and mass spectrometry" (IP/LC–MS, short IP/MS) is widely used to detect proteins and their interactions in a cell sample. Immunoprecipitation allows the selective extraction of a known protein from the sample together with other proteins that are currently bound to it. The extracted assembly can be analyzed with liquid chromatography and mass spectrometry to discover and identify the interaction partners of the known protein.

Proteomics is a research field within biology that focuses on detecting and quantifying proteins and their interactions. IP/MS is a typical proteomics laboratory tool, because IP/MS can be performed cell–wide to identify the interaction partners of all known proteins in a cell. By combining all the results, we can build a dataset that stores information about which protein was extracted together with which protein from the sample. This dataset can be further analyzed using computational tools to detect protein complexes in the cell.

Computational complex–finding tools are available since many years, but it is still difficult to determine the interaction level of an identified protein complex,

i. e., whether it is a complex–core or a larger assembly built of several cores. Considering these different protein interaction levels has caused a shift of thinking in proteomics, called the transition from proteomics to complexomics. The focus has changed from treating all protein interactions equally to a subtle distinction between binary interactions, stable complexes, and larger assemblies.

This thesis follows the complexomics way of thinking and combines my complexome research of the last four years. It introduces novel methods for finding and visualizing protein complexes in IP/MS data. The methods are designed as interactive algorithm-based and rule-based tools, which help users detect protein complexes in different IP/MS datasets, visualize them and obtain information about their interaction level.

Chapter 1 introduces the field of IP/MS-based proteomics and explains how the different protein interaction levels can influence the IP/MS data. Chapter 2 presents a new computational method that uses simple rules to find protein complexes and to detect their interaction level. This chapter also introduces a graphical tool for visualizing protein complexes from large–scale IP/MS results. Chapter 3 describes a comprehensive extension of the previously introduced method, which can automatically capture protein complexes at different interaction levels at once. Extended visualization tools that are able to display complexes and their interaction levels in different ways are also introduced in this chapter. Chapter 4 presents a strategy to predict higher protein interaction levels from IP/MS data where the complex–cores are already known.