

Abstract

Protein-protein interactions (PPI) play key roles in determining the outcome of most cellular processes. Correctly identifying and characterizing protein interactions and the networks they comprise is critical for understanding the molecular mechanisms of many biological functions carried out by the cell. Computational approaches could be utilized to combine multiple information sources in order to predict the sets of interacting protein pairs and to identify important biological substructures in this network.

In this dissertation, a systematic study is first carried out to evaluate the efficacy of using supervised learning methods to integrate direct and indirect biological evidence for predicting pairwise protein interactions. Then, four learning algorithms are developed to derive PPI networks from different perspectives.

(I) A combined computational and experimental approach is proposed for predicting interaction partners of human membrane receptors.

(II) Considering that no large set of non-interacting protein pairs exists, we employ a ranking approach to identify potential interaction pairs that are "similar" to known interacting pairs.

(III) A multiple-view learning strategy (referred to as "mixture of feature experts") is derived for predicting PPIs that takes into account the heterogeneous nature of features. The derived weighting of the experts (feature groups) provides a way to evaluate each prediction based on its high scoring feature experts.

(IV) "Protein complex" (a special group formation) is one typical pattern contained in protein-protein interaction networks. We present an algorithm for inferring protein complexes based on graph topological patterns and biological properties with supervised graph clustering.

Our proposed algorithms provide strong computational tools for predicting and analyzing protein-protein interaction networks. They have been applied successfully in yeast and human, and have generated promising results. For instance, without the novel interaction between rhodopsin and chemokines found by our computational approach, the important functional implication of rhodopsin in the immune system would not have been possibly discovered.