Chapter 12
Computational Methods for Prediction of Protein-Protein Interactions:
PPI Prediction Methods

Sneha Rai
Netaji Subhas Institute of Technology, India

Sonika Bhatnagar
Netaji Subhas Institute of Technology, India

ABSTRACT

The key signaling pathways in cellular processes involve protein-protein interactions (PPIs). A perturbation in the balance of PPIs occurs in various pathophysiological processes. There are a large numbers of experimental methods for detection of PPIs. However, experimental PPI determination is time consuming, expensive, error prone and does not effectively cover transient interactions. Therefore, overlaying and integration of predictive methods with experimental results provides statistical robustness and biological significance to the PPI data. In this chapter, the authors describe PPIs in terms of types, importance, and biological consequences. This chapter also provides a comprehensive description on various computational approaches for PPI prediction. Prediction of PPI can be done through: 1) Genomic information based methods 2) Structure based methods 3) Network topology based methods: 4) Literature and data mining based methods 5) Machine learning methods. For ease of use and convenience, a summary of various databases and software for PPI prediction has been provided.

INTRODUCTION

Proteins are essential macromolecules that perform diverse functions in association with other macro and small molecules. Biological processes involve interplay between protein-protein interactions (PPIs) as cell to cell interactions, metabolic and developmental processes are governed by PPIs. A perturbation in the balance of PPIs occurs in various pathophysiological processes due to varying expression

DOI: 10.4018/978-1-5225-1798-6.ch012

levels of disease related proteins. Specifically, PPIs also modulate the host-pathogen interactions in the initial stages and host response in later stages of infection. Modeling PPIs at a large scale for study of topological parameters and essentiality leads to identification of crucial players and pathways in the disease condition. Since proteins may have different interacting partners under diseased and normal state respectively, identification of drug targets from PPI networks is also carried out. Targeting PPIs presents a novel approach for developing therapeutics with high specificity and fewer side effects. Moreover, PPIs aid in forecasting the function of a target protein and drugability of a molecule. There are a large numbers of experimental methods for detection of PPIs that have led to the accumulation of PPI data in publicly available databases. However, experimental PPI determination is time consuming, expensive, error prone and do not effectively cover transient interactions. Therefore, prediction of PPIs by computational methods represents an important development. Overlaying and integration of predictive methods with experimental results provides statistical robustness and biological significance to the PPI data.

The chapter provides an overview on PPIs, its types, importance, biological consequences and factors affecting PPIs. The authors have also provided a comprehensive description on various computational approaches for PPI prediction. Computational methods for PPI prediction are based on sequence, structure, evolution and genomic data. Prediction of PPI can be done through:

1. Genomic information based methods,
2. Structure based methods,
3. Network topology based methods,
4. Machine learning methods and
5. Literature and data mining based methods.

Increment in genetic information obtained through complete genome sequencing methods, had laid down the foundation for PPI prediction at genetic level. These methods are based on the concept that genes that functionally interact are often located in close proximity in the genome. Proteins that interact they co-adapt to the evolutionary forces in order to preserve the interactions. Genomic context based methods apply the concept of genetic linkage, gene fusion, phylogenetic profiles and *in silico* two hybrid systems for PPI prediction. Some computational methods utilize three dimensional information of the protein to predict PPI. The fact behind these methods is that interactions between two proteins can be predicted if interactions between their homologous proteins are known. In fact, interactions between analogous proteins can also be predicted by fold recognition or threading based methods. Protein-protein docking methods are also employed for forecasting interactions between proteins as these resolve the native structure of the protein complexes. Prior knowledge about the interacting residues often reduces the search space for the most optimal complex. Most of the docking methods consider proteins as rigid bodies and consider shape complementarities between interacting surfaces as the main guide. Only a few approaches take protein flexibility in consideration where the interacting proteins are allowed to adopt the active bound conformation. These methods are limited to structural information of a protein but give a more accurate view to PPIs. Biological networks, where nodes represent proteins and edge represents interaction between them, possesses certain characteristics that differentiate them from random networks. In network based methods for PPI prediction, a confidence score is assigned to each edge of the network that facilitate distinction between protein pairs that actually interact from protein pairs that exist by chance. Usually, the confidence score for a particular interaction is calculated by comparing original networks with random networks which are formed based on graph and probability theory. PPI