Computational Methods for Maximizing Confidence and Significance in Protein-Protein Interactions

Sneha Rai[1] and Sonika Bhatnagar[2]

[1,2]Computational and Structural Biology Laboratory, Division of Biotechnology, Netaji Subhas Institute of Technology, Dwarka, New Delhi–110078

Abstract—Detection of Protein-Protein Interactions (PPIs) is necessary for understanding the molecular basis of biological processes, using network biology approach for identifying central payers in physiological and pathological phenomena, and for targeting PPIs with therapeutic agents. As the interacting partners of a protein vary under diseased and normal state, targeting PPIs is a novel approach for developing new therapeutics with fewer side effects. There are large numbers of experimental methods for PPI detection and have led to development of PPI databases. However, these methods are expensive, time consuming, error prone and often unable to detect transient interactions. Thus, prediction of PPIs through computational methods represents an important advancement in maximizing the statistical significance and biological relevance of experimental methods. Integration of experimental and computational data increases the robustness and biological significance of PPIs. In this work, we present a summary of the computational resources and methods for PPI prediction including a) Genomic context based methods, which utilize the concepts of gene fusion, gene co-localization and conservation of function to predict PPIs. b) Structural complementarity based methods use three dimensional information of the proteins to forecast interactions between them. c) Network topology based methods where the underlying data may be mined from Literature or experimental techniques. Experimentally determined high confidence interactions are then used to construct and visualize networks. Network based statistics and parameters lend completeness, clarity and visual overview for computing choke points in the system; d) Machine learning methods combining multiple parameters. Machine learning methods construct a model using training set with known outcomes, which is then used to predict interacting proteins. Artificial neural networks, support vector machine, random forest, decision tree, naïve bayes and k-nearest neighbors are machine learning based methods currently in use for PPI prediction.