
IGI PUBLISHING **ITB15229**
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.igi-global.com

This paper appears in the publication, **Data Mining and Knowledge Discovery Technologies** edited by D. Taniar © 2008, IGI Global

Chapter XII

Study of Protein-Protein Interactions from Multiple Data Sources

Tu Bao Ho
Japan Advanced Institute of Science and Technology, Japan

Thanh Phuong Nguyen
Japan Advanced Institute of Science and Technology, Japan

Tuan Nam Tran
Japan Advanced Institute of Science and Technology, Japan

Abstract

The objective of this chapter is twofold. First is to provide a survey of computational methods for protein-protein interaction (PPI) study. Second is to introduce our work and results in using inductive logic programming to learn prediction rules for PPI and DDI (domain-domain interactions) from multiple data sources. We show advantages of exploiting various types of data in these important problems of bioinformatics.

Introduction

Proteins are macromolecules made of 20 amino acids arranged in a linear chain, which participate in every process within cells. Many proteins play a key role in biochemical reactions, and structural or mechanical functions, and thus understanding functions of proteins is a main task in molecular biology. Early work has focused on finding protein functions via prediction of protein structures (Bock & Gough, 2001; Marcotte et al., 1999). Recently, detecting protein functions via protein-protein interactions (PPI) has emerged as a new trend in computational biology (Baudot et al., 2006; Chen & Liu, 2005; Chen & Yuan, 2006). Protein-protein interaction study is not only crucial in finding protein functions, but also is a significant task, as protein interactions are one of the most important regulatory mechanisms in cells, and most of the cellular processes are coordinated by specific protein interactions. For example, from the physical association between a novel protein and a well-characterized protein, we can infer the functions of the former.

Discovering protein-protein interactions has been a key problem in molecular biology and bioinformatics. Some good surveys about protein-protein interaction research have been available (Ng & Tan, 2004; Uetz & Vollert, 2006). Generally, there are experimental and computational methods for prediction of protein interactions. The experimental methods are divided into two groups, the traditional and the high-throughput ones. Traditional experimental methods typically include co-immunoprecipitation and synthetic lethal screening. Although the high-throughput experimental detection methods for PPI (typically, yeast two-hybrid, phage display, affinity purification and mass spectrometry, and protein micro-arrays) present many advantages over traditional experimental methods, but they are still tedious, labor-intensive, and usually have high false positive and high false negative rates.

Computational methods for detecting protein interactions, recently developed with various machine learning techniques and various types of available biological data, allow a chance to study more widely and deeply about protein-protein interactions.

The objective of this chapter is twofold. First is to briefly review the key ideas, advantages, and limitations of computational methods for PPI prediction, some new trends, and potential usage of PPI prediction results. Second is to present our work on PPI and DDI prediction using inductive logic programming on multiple sources of genomic and proteomic data.

An Overview of Computational Methods for PPI Prediction

The main target of computational methods for PPI prediction is to learn patterns or models from available genomic/proteomic data, which can be used to predict the interaction between given proteins.