

# Chapter XII

## Predicting Protein Functions from Protein Interaction Networks

**Hon Nian Chua**

*Institute for Infocomm Research, Singapore*

**Limsoon Wong**

*National University of Singapore, Singapore*

### **ABSTRACT**

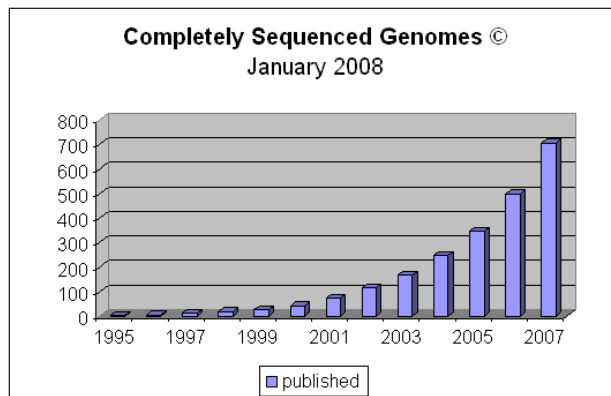
*Functional characterization of genes and their protein products is essential to biological and clinical research. Yet, there is still no reliable way of assigning functional annotations to proteins in a high-throughput manner. In this chapter, the authors provide an introduction to the task of automated protein function prediction. They discuss about the motivation for automated protein function prediction, the challenges faced in this task, as well as some approaches that are currently available. In particular, they take a closer look at methods that use protein-protein interaction for protein function prediction, elaborating on their underlying techniques and assumptions, as well as their strengths and limitations.*

### **INTRODUCTION**

Since the completion of the Human Genome Project (HGP) in 2003, genomic and proteomic research has gained much momentum. Based on statistics from the Genome OnLine Database (GOLD) (Liolios et al. 2008), the number of genomes sequenced grew exponentially since 1995, with nearly 700 genomes completely sequenced by 2007 (See Figure 1). With the maturation of genomic data generation, the focus in biological research has shifted towards the understanding of the complex functional and interactive processes between proteins and multi-component molecular machines that contribute to the majority of operations in cells, as well as the transcriptional regulatory mechanisms and pathways that control these cellular processes (Frazier et al. 2003). There is also a pressing need for the functional characterization of genes in clinical research to better understand diseases (Hu et al. 2007).

In contrast to the unprecedented rate at which new genes are being discovered, the pace at which novel genes and their corresponding protein products are characterized pales in comparison. A recent survey on function

Figure 1. Number of completely sequenced genomes from year 2005 to 2007. Credit: Image adapted from [http://www.genomesonline.org/gold\\_statistics.htm](http://www.genomesonline.org/gold_statistics.htm).



prediction techniques showed that out of 345 genomes listed in the KEGG Genome collection (Kanehisa et al. 2004), 222 have some ambiguous functional annotations assigned to half or more of its genes (putative, probable, and unknown) (Hawkins et al. 2007). This may be attributed to the lack of reliable high-throughput method to identify the functional nature of proteins. Unlike genomic sequences, function is an abstract and complex notion, and can only be ascertained through the observation of multiple aspects of a protein, such as its sequence, structure, interaction behavior and changes in phenotype upon its mutation or removal.

Besides the influx of genomic sequence data, the maturation of high-throughput techniques for various other genomic analyses such as gene expression profiling (Eisen et al. 1998; Hughes et al. 2000), immuno-precipitation, genetic interactions, two-hybrid (Gietz et al. 1997), tandem-affinity purification, mass spectrometry, and more recently, flow cytometry and Protein-Fragment Complementation Assay (Tarassov et al. 2008), also makes available a wealth of other biological data. Advancements in computational techniques such as secondary and tertiary structure prediction also make it possible to generate computationally predicted data in large scale (Rost et al. 2003). This multitude of heterogeneous information presents to researchers a global perspective of the mechanisms behind genes and their protein products, and offers hope to elucidate the functions of proteins which cannot be easily characterized by sequence alone. However, this escalating rate of growth in biological data also makes manual annotation of protein function an increasingly daunting task. This paves the way to the emergence and popularization of automated function prediction. While it is unlikely that automated function prediction can produce authoritative annotations, it can provide systematic identification of potential novel annotations, which may be used to guide the prioritization of resource allocation for experimental verification. This can potentially improve the throughput of conventional functional characterization.

## Objectives of the Chapter

In this chapter, we hope to provide a concise overview on the background, challenges and approaches taken in the task of automated protein function prediction, especially for methods that utilize protein-protein interactions. The chapter is organized as follows: First, we provide an overview of techniques used in automated protein function prediction. Second, we discuss the difficulties that are associated with this task. Finally, we describe in greater detail some of the techniques that use protein-protein interactions to predict protein function.

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/predicting-protein-functions-protein-interaction/5566](http://www.igi-global.com/chapter/predicting-protein-functions-protein-interaction/5566)

## Related Content

---

### Domain-Based Approaches to Prediction and Analysis of Protein-Protein Interactions

Morihiro Hayashida and Tatsuya Akutsu (2014). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 24-41).

[www.irma-international.org/article/domain-based-approaches-to-prediction-and-analysis-of-protein-protein-interactions/105099](http://www.irma-international.org/article/domain-based-approaches-to-prediction-and-analysis-of-protein-protein-interactions/105099)

### Local Optima Avoidance in GA Biclustering using Map Reduce

Gowri R. and Rathipriya R. (2016). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 37-47).

[www.irma-international.org/article/local-optima-avoidance-in-ga-biclustering-using-map-reduce/171418](http://www.irma-international.org/article/local-optima-avoidance-in-ga-biclustering-using-map-reduce/171418)

### Proteomics in Personalized Medicine: An Evolution Not a Revolution

Srijan Goswami (2018). *Applying Big Data Analytics in Bioinformatics and Medicine* (pp. 80-98).

[www.irma-international.org/chapter/proteomics-in-personalized-medicine/182942](http://www.irma-international.org/chapter/proteomics-in-personalized-medicine/182942)

### Incorporating Network Topology Improves Prediction of Protein Interaction Networks from Transcriptomic Data

Peter E. Larsen, Frank Collart and Yang Dai (2010). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 1-19).

[www.irma-international.org/article/incorporating-network-topology-improves-prediction/47093](http://www.irma-international.org/article/incorporating-network-topology-improves-prediction/47093)

### Big Data Visualization for Black Sigatoka Disease of Bananas and Pathogen–Host Interactions (PHI) of Other Plants

Richard S. Segall, Soichiro Takahashi and Prasanna Rajbhandari (2024). *International Journal of Applied Research in Bioinformatics* (pp. 1-22).

[www.irma-international.org/article/big-data-visualization-for-black-sigatoka-disease-of-bananas-and-pathogen-host-interactions-phi-of-other-plants/361940](http://www.irma-international.org/article/big-data-visualization-for-black-sigatoka-disease-of-bananas-and-pathogen-host-interactions-phi-of-other-plants/361940)