

Chapter 13

Mining Protein Interactome Networks to Measure Interaction Reliability and Select Hub Proteins

Young-Rae Cho
Baylor University, USA

Aidong Zhang
State University of New York, Buffalo, USA

ABSTRACT

High-throughput techniques involve large-scale detection of protein-protein interactions. This interaction data set from the genome-scale perspective is structured into an interactome network. Since the interaction evidence represents functional linkage, various graph-theoretic computational approaches have been applied to the interactome networks for functional characterization. However, this data is generally unreliable, and the typical genome-wide interactome networks have a complex connectivity. In this paper, the authors explore systematic analysis of protein interactome networks, and propose a k -round signal flow simulation algorithm to measure interaction reliability from connection patterns of the interactome networks. This algorithm quantitatively characterizes functional links between proteins by simulating the propagation of information signals through complex connections. In this regard, the algorithm efficiently estimates the strength of alternative paths for each interaction. The authors also present an algorithm for mining the complex interactome network structure. The algorithm restructures the network by hierarchical ordering of nodes, and this structure re-formatting process reveals hub proteins in the interactome networks. This paper demonstrates that two rounds of simulation accurately scores interaction reliability in terms of ontological correlation and functional consistency. Finally, the authors validate that the selected structural hubs represent functional core proteins.

DOI: 10.4018/978-1-4666-1785-8.ch013

INTRODUCTION

Understanding functional behaviors of molecular components is an underlying base for biomedical applications. A wide range of computational approaches have been applied to characterize molecular functions from various types of data sources. In the past, sequence or structure analysis of proteins has contributed to characterize their functions. However, they are incapable of systematically analyzing complex functional mechanisms through biochemical reactions or interactions. Proteins typically execute their functions through interactions with other biomolecular units. Comprehensive knowledge of protein-protein interactions is thus essential to understanding the intrinsic mechanisms of biological processes.

Earlier data of protein-protein interactions were obtained via intensive small-scale investigations of restricted sets of proteins of interest, each yielding data sets regarding a limited number of protein-protein interactions. However, recent high-throughput techniques, such as yeast two-hybrid systems and mass spectrometry, involve genome-wide detection of protein-protein interactions (Uetz et al., 2000; Ito, Chiba, Ozawa, Yoshida, Hattori, & Sakaki, 2001; Gavin et al., 2002; Ho et al., 2002; Giot et al., 2003; Li et al., 2004). The yeast two-hybrid system (Parrish, Gulyas, & Finley, 2006) seeks feasible binary interactions between any two proteins encoded in the genome of interest. The interaction of two proteins transcriptionally activates a reporter gene. This reaction tracks the interaction, revealing “prey” proteins that interact with a known “bait” protein. The mass spectrometry (Aebersold & Mann, 2003) analyzes the composition of a partially purified protein complex. It uses an affinity tag attached to target “bait” proteins for purifying complexes. Comprehensive protein-protein interaction data sets in model organisms, generated by the high-throughput experiments, are publicly available in a number of open databases such as BioGRID (Breitkreutz, Stark, Reguly, Boucher,

Breitkreutz, Livstone, Oughtred, Lackner, Bahler, Wood, Dolinski, & Tyers, 2008), MIPS (Mewes, Dietmann, Frishman, Gregory, Mannhaupt, Mayer, Munsterkötter, Ruepp, Spannagl, Stumppf, & Rattei, 2008), DIP (Salwinski, Miller, Smith, Pettit, Bowie, & Eisenberg, 2004), MINT (Chatr-aryamontri, Ceol, Montecchi-Palazzi, Nardelli, Schneider, Castagnoli, & Cesareni, 2007), IntAct (Aranda et al., 2010), and HPRD (Prasad et al., 2009). However, accurate analysis of protein-protein interactions has been limited due to unreliable interaction data. The large-scale experimental data sets are susceptible to false positives, i.e., some fraction of the putative interactions detected should be considered spurious because they cannot be confirmed to occur *in vivo* (von Mering, Krause, Snel, Cornell, Oliver, Fields, & Bork, 2002; Sprinzak, Sattath, & Margalit, 2003).

The protein-protein interaction data sets from a genome-scale perspective are structured into an interactome network, which is represented as an undirected, un-weighted graph having proteins as a set of nodes and interactions between them as a set of edges. Specific topological features have been observed in the networks, such as scale-free distribution (Barabasi & Albert, 1999), small-world property (Watts & Strogatz, 1998), and hierarchical modularity (Ravasz, Somera, Mongru, Oltvai, & Barabasi, 2002). Various graph-theoretic computational approaches (Sharan, Ulitsky, & Shamir, 2007) have been applied to the interactome networks for functional characterization. Connection pattern is a major attribute to predict functions of unknown proteins in the networks (Oliver, 2000). However, accurate analysis of interactome networks has been challenged because of complex connectivity. In general, this complexity is caused by multiple functions of each protein and numerous cross-links between functional modules (Hartwell, Hopfield, Leibler, & Murray, 1999). A functional module is defined as a group of proteins that participate in the same functional activities (Spirin & Mirny, 2003). When handling such complex interactome networks, classical graph-theoretic

15 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/mining-protein-interactome-networks-measure/66713

Related Content

Work Time Allocation of Family Doctors in Rural Tajikistan

Nargis S. Juraeva, Martin W. Bratschi, Kaspar Wyss and Ismoil S. Komilov (2020). *International Journal of Applied Research in Bioinformatics* (pp. 28-37).

www.irma-international.org/article/work-time-allocation-of-family-doctors-in-rural-tajikistan/261868

Prioritizing Disease Genes and Understanding Disease Pathways

Xiaoyue Zhao, Lilia M. Iakoucheva and Michael Q. Zhang (2012). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 31-49).

www.irma-international.org/article/prioritizing-disease-genes-and-understanding-disease-pathways/101241

Introduction: The Landscape of Pulmonary Bioinformatics

Rahul Kumar, Devvret Verma, Bahman Khoshru and Adeyemi Nurudeen Olatunbosun (2026). *Next-Generation Bioinformatics for Pulmonary Disease Research* (pp. 1-36).

www.irma-international.org/chapter/introduction/393257

Can Activated Platelet Rich Plasma Combined with Adipose-Derived Stem Cells Be Used to Treat Skin Wrinkles?: A Mechanism Study

Phuc Van Pham, Loan Thi-Tung Dang, Nhung Hai Truong and Ngoc Kim Phan (2013). *Bioinformatics: Concepts, Methodologies, Tools, and Applications* (pp. 920-936).

www.irma-international.org/chapter/can-activated-platelet-rich-plasma/76102

Homology Modeling and Evaluation of Sars-Cov-2 Spike Protein Mutant: D614G

Hima Vyshnavi, Aswin Mohan, Shahanas Naisam, Suvanish Kumar and Nidhin Sreekumar (2024). *Research Anthology on Bioinformatics, Genomics, and Computational Biology* (pp. 948-967).

www.irma-international.org/chapter/homology-modeling-evaluation-sars-cov/342557