Chapter I
Molecular Biology of Protein–Protein Interactions for Computer Scientists

Christian Schönbach
Nanyang Technological University, Singapore

ABSTRACT

Advances in protein-protein interaction (PPI) detection technology and computational analysis methods have produced numerous PPI networks, whose completeness appears to depend on the extent of data derived from different PPI assay methods and the complexity of the studied organism. Despite the partial nature of human PPI networks, computational data integration and analyses helped to elucidate new interactions and disease pathways. The success of computational analyses considerably depends on PPI data understanding. Exploration of the data and verification of their quality requires basic knowledge of the molecular biology of PPIs and familiarity with the assay methods used to detect PPIs. Both topics are reviewed in this chapter. After introducing various types of PPIs the principles of selected PPI assays are explained and their limitations discussed. Case studies of the Wnt signaling pathway and splice regulation demonstrate some of the challenges and opportunities that arise from assaying and analyzing PPIs. The chapter is concluded with an extrapolation to human systems biology that offers a glimpse into the future of PPI networks.

INTRODUCTION

The function and functioning of a cell from differentiation to death depends on information flow among its molecular and cellular components. Processes enabling intra- and extracellular information flow include PPIs, signal transduction, and protein transport, among others. Methods to identify PPIs (reviewed by Berggård et al., 2007; Lalonde et al., 2008) *in vitro a*nd *in vivo* have produced thousands of interaction pairs that are presented in mostly static interactomes. For the most part, full understanding of the biological significance of interactions in cellular context, particularly in human and mammalian model organisms remains a challenge. Owing to the intrinsic properties of PPI assays none of the interactomes are complete. The best sampled and curated (Reguly et al., 2006) interactome of yeast is thought to represent about half of the complete PPI set (Hart et al., 2006).

Each PPI assay has different biases for soluble or membrane proteins, protein abundance, affinity range, and post-translational modifications that define stable or transient interactions (Charbonnier et al., 2008). Coordination and standardization (Hermjakob et al., 2004; Kerrien et al., 2007) of protein interaction data has improved analyses and integration, but a reference human interactome that can be used for benchmarking is not yet available. In the coming five years results of the Human Interactome Project (2008) and the Human Proteome Organization (HUPO) initiative to map the human proteome (Pearson, 2008) are expected to improve the molecular network-based understanding of biological processes and human diseases. However, successful development of diagnostics and therapeutics that target disease-relevant PPIs will depend on the progress of obtaining quantitative and dynamic PPI data across different tissue cells and their integration with gene expression, structural, and metabolic data. In view of the following chapters this review will give a brief tour of PPI interaction types and detection methods illustrated by biological interesting examples. Caveats and aspects of biological properties that are of interest for computational analyses and predictions of PPIs are highlighted.

BACKGROUND: PPI TYPES

PPI as any other molecular interaction must meet at least two conditions: sufficient concentration of interactants and a favorable difference of free energy upon interaction. Both parameters are dynamic and depend on the cellular environment. The concentration of a protein may depend on the level of gene expression, translation efficacy, its stability or turnover, and availability in the same cellular compartment as the potential interacting partner. The level of free of energy difference depends on the sum free energy differences derived from atomic level interactions such as hydrogen bond formation, van der Waals interactions and ionic bond formation. Corresponding to the difference of total free energy the resulting interactions are broadly classified into stable, strong transient, or weak transient interactions. Stable interactions are often formed by homodimers, for example RNA polymerase II. Many heterodimeric interactions, for instance the binding of cell division cycle 42 to Wiskott-Aldrich syndrome protein (Rudolph et al., 1998) are transient, but relatively strong. Interactions of cell surface receptors or cargo proteins are weak and transient (e.g., carbohydrate mediated binding of lectin mannose binding 1 to cathepsin C (Nyfeler et al., 2005)).

The detection of PPIs depends on the intrinsic properties and sensitivities of the assays which are reviewed in detail by Berggård (2007) and Lalonde (2008). Briefly, medium throughput surface plasmon resonance assays (e.g., BIACore™) are among the most sensitive *in vitro* assay systems that capture also the kinetics of PPIs. The popular *in vivo* yeast two-hybrid (Y2H) assay is high throughput, but not quantitative and biased towards the detection of stronger binary interactions between protein domains.

PPI DETECTION

Principles of Y2H Method

Appropriate use of wet lab data in computational analyses requires at least a basic understanding of the method(s) used to generate the data. Here, the principles of two commonly used large-scale PPI detection methods, Y2H and tandem affinity purification-mass spectrometry (TAP-MS) are introduced and their limitations discussed.

Almost 20 years ago Field and Song (1989) published a genetic system to detect PPIs that is now termed Y2H. The authors utilized the transcriptional activation properties of GAL4, which depend on a DNA binding domain (DBD) and an activation domain (AD). GAL4 binds with its DBD to specific upstream sites (UAS promoter) of genes encoding enzymes of the galactose metabolism (Fig. 1A). At the same time, AD facilitates the transcription of the downstream genes by RNA polymerase II. When DBD and AD were physically separated transcriptional activation was abolished, whereas reconstitution of DBD and AD enabled transcription. Indeed, when Fields and Song fused DBD to SNF1 and AD to SNF4, the interaction of SNF1 and SNF4 reconstituted the transcriptional activity of GAL4. In Y2H assays (Fig. 1B) the hybrid protein DBD-X is called bait. The fusion of AD with