Protein protein interaction network

Durga Rai 1, Mitrabinda Kheti2, Madalasa Kheti 3, Mohit Kamthania4
1,2,4 Department of Biotechnology, Faculty of Life Science, Institute of applied medicines and research, Ghaziabad, India.
3 School of basic and applied Sciences, Galgotias University, Greater Noida, India.

Abstract - Protein-protein interactions (PPIs) and multi-protein complexes perform central roles in the cellular systems of all living organisms. In humans, disruptions of the normal patterns of PPIs and protein complexes can be causative or indicative of a disease state. Recent developments in the biological applications of mass spectrometry (MS)-based proteomics have expanded the horizon for the application of systematic large-scale mapping of physical interactions to probe disease mechanisms. In this review, we examine the application of MS-based approaches for the experimental analysis of PPI networks and protein complexes, focusing on the different model systems (including human cells) used to study the molecular basis of common diseases such as cancer, cardiomyopathies, diabetes, microbial infections, and genetic and neurodegenerative disorders.

Index Terms - PPI, Complex diseases, Networks

INTRODUCTION

Early biological experiments revealed that proteins, as the main agents of biological function, determine the phenotype of all organisms. By the advent of molecular biology, it has been assumed that proteins are not naturally functional in isolated forms; instead, they have interactions with one another and also with other molecules (e.g. DNA, RNA) that mediate metabolic and signalling pathways, cellular processes, and organismal systems (1). Thus, studies of proteins’ interactions are fundamental to perceive their role within the cell. The term ‘protein interaction’ encompasses a variety of events, such as transient and stable complexes, as well as physical and functional interactions (2). Protein-protein interaction (PPI) data can be used in a larger scale to map networks of interactions depend on their physical or functional association (3, 4). Protein interaction networks are practical means to abstract basic knowledge and to improve biological and biomedical applications. Although protein interaction networks are incomplete (5) and error-prone (6), systematic studies of them have been confirmed to be especially important for deciphering the relationships between network structure and function (7), discovering novel protein function (8), identifying functionally coherent modules (9, 10), and conserved molecular interaction patterns (11, 12). Since proteins have principle role in biological function, their interactions determine molecular and cellular mechanisms which control healthy and diseased states in organisms. Diseases are often caused by mutations affecting the binding interface or leading to biochemically dysfunctional allosteric changes in proteins (13). Therefore, the molecular basis of diseases can be enlightened through protein interaction networks, which in turn can appraise methods for prevention, diagnosis, and treatment. Generally, traditional analyses exploit a univariate approach to study gene expression and identify genes with meaningful individual differential expression in the phenotype of interest (14). However, the underlying mechanisms of complex diseases, which arise from the interplay among multiple genetic and environmental factors, cannot be explicated by such univariate approaches. Hence, since there are remarkable increase in an availability of human protein interaction data, the focus of bioinformatics development has shifted from understanding networks encoded by model species to understand the networks underlying human disease (15).
Protein-protein interactions (PPIs) are central to the proper functioning of the most basic molecular mechanisms underlying cellular life and are often perturbed in disease states. It is predicted that the human complement of PPIs (the interactome) numbers between 130,000 and 600,000 [1, 2]. These include interactions of structural proteins inside the cell, and multi-protein complexes that are involved in core processes such as transcription and translation, cell-cell adhesion and communication, protein synthesis and degradation, cell cycle control and signalling cascades. The study of PPI networks and the global physical organization of cells is needed to provide a better understanding of basic cellular biochemistry and physiology (Figure 1). It is therefore no surprise that when the homeostatic state of an organism or an individual cell is disturbed (as a result of environmental stress or in a disease state) the 'normal' patterns of PPIs are disturbed. Many of these disruptions can often be considered side products of a disease that have no significant functional consequence, but others can frequently play a major causal role in disease and have a central impact on the initiation or progression of a pathology (Figure 1). For example, the role of PPI disturbances in the interactome of the p53 tumour suppressor protein, caused by mutations in its gene, are well established [3, 4]: disruptions in the desmosome-mediated interactions between cells have been implicated in a variety of diseases [5]; aberrant PPIs causing the accumulation of protein aggregates can result in a number of neurodegenerative diseases [6, 7]; and host-pathogen PPIs are of central importance in infection [8, 9]. Therefore, depending on the pathological scenario, the monitoring and study of PPIs in different biological models can provide interesting and significant options for both diagnostic and therapeutic targets that have potential for broader clinical applicability. The major biomedical goal of identifying and studying PPI networks in disease states is the development of therapies targeting interactions that are functionally relevant to disease progression and patient outcomes. Another long-term clinical goal would be the identification of disease-specific patterns of PPIs, which could serve as disease- or treatment-responsive biomarkers whose selective measurement leads to improved diagnostics or prognostics for common human disorders.

Technological advances in genomics and proteomics have spawned a large number of comprehensive studies that, in turn, have generated huge amounts of data. In recent years, innovative developments in the application of highly sensitive and accurate forms of mass spectrometry (MS) to biological specimens have provided considerable progress in the rapidly emerging fields of metabolomics, lipidomic, glycemic and proteomics. These include the large-scale identification and characterization of a number of post-translational modifications (PTMs) on proteins (phosphorylation, glycosylation, ubiquitylation, methylation and so on). Most notable, however, advances in large-scale protein-interaction mapping have led to a significant expansion in our understanding of both the composition of protein complexes and their arrangement within broader cellular PPI networks that are often perturbed under disease states. There have been several reviews of technical developments in the identification and characterization of PPIs and protein complexes [10–13]. Here, we examine the application of MS-based experimental analyses of model systems to explore heterogeneous PPI networks and protein complexes in the context of human disease.

MICROBES AS CELL MODELS

Unicellular organisms such as yeast have served as tractable models to probe the molecular biology of eukaryotes, whereas most major human pathogens are prokaryotes. Hence, PPIs have been studied in microbes in great detail. Several landmark studies have contributed greatly to our understanding of the role PPI networks play at all levels of life. The first
studies utilizing MS-based approaches in studying PPIs were performed in two of the most basic model systems used in molecular biology, the Gram-negative bacterium Escherichia coli and the budding yeast Saccharomyces cerevisiae. Owing to their experimental amenability (in terms of genetic manipulation, generation time and so on), these model systems have proven invaluable in proof-of-concept method development in the MS-based interactomics field. Important from a clinical perspective, a significant number of complexes and PPIs that have been mapped in microbes are conserved (to a varying degree) in humans, and disturbances in their normal homeostatic patterns can be indicative or even causative in disease conditions.

The most suitable methodology for study of protein complexes and PPI in these model systems has proven to be the affinity purification of protein complexes followed by MS identification (AP-MS). The existence of genome-scale libraries of genetically engineered E. coli and yeast strains expressing individually tagged proteins from native promoters has allowed for the relatively rapid isolation and large-scale mapping of stable protein interactomes in both of these organisms, including most recently membrane-associated complexes [21]. Tandem affinity purification (TAP) [22, 23] and sequential peptide affinity (SPA) tagging technologies [24, 25] have also contributed to the streamlining of AP-MS identification and characterization of PPIs and heterogeneous protein complexes. These methods allowed for the unprecedented characterization of widely conserved protein complexes in yeast [26] and E. coli [27].

Because they are eukaryotic and show a greater degree of conservation with humans, baker’s yeast has been a particularly informative model of human protein complexes and PPIs. Several landmark studies have utilized AP-MS to map the yeast protein interactome in a comprehensive manner [28–33]. Two of the more comprehensive studies, from our group and that of a competing company (Cellzome), applied matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) and liquid chromatography (LC)-MS in association with TAP of more than 4,500 tagged yeast proteins to map more than 7,000 interactions and to identify 429 putative protein complexes [26, 34]. Notable aspects of the two studies were the high technical reproducibility and the reciprocal tagging and purification of candidate interactors that provided an estimate of reliability. Strikingly, however, despite using a similarly stringent experimental approach and being co-published at the same time, the overlap of the predicted complexes and PPI was initially found to be low. This discrepancy was widely interpreted as suggesting the incompleteness or unreliability of high-throughput interaction data, but it was later ascribed to differences in the computational scoring and post-processing of each PPI network, indicating that inconsistent data analysis is a major outstanding issue for the field. In a more recent follow-up study in yeast by our group, a carefully defined set of 501 heterogeneous membrane protein complexes were charted in yeast through the additional analysis and identification of detergent-solubilized proteins [21]. A protein kinase-phosphatase interaction network encompassing transient dynamic regulator-substrate interactions has also been mapped using a modified AP-MS-based approach [35].

Owing to the requirement for novel therapeutics and the related need for understanding molecular pathogenesis, PPIs involving pathogenic bacteria and viruses have also garnered significant attention. In the study of viruses, the major focus is the discovery of novel protein-based antigens for the development of vaccines. The mechanisms of host-pathogen interactions and how the pathogen co-opts the host's molecular machinery have also been studied through the examination of host-pathogen PPIs [8]. MS-based methodologies for virus-host proteomics have been reviewed recently [9]. As a result of recent studies of the HIV interactome, several host and viral proteins have been discovered to play a crucial role in the life cycle of infection and appear to have provided potential novel therapeutic targets. An extensive AP-MS-based study of the HIV host-pathogen PPIs was performed [36] by expression of individual tagged HIV proteins transiently in the human embryonic kidney 293 (HEK293) cell line or stably in Jurkat cells (immortalized T lymphocytes) [37]. Putative PPIs from AP-MS were confirmed by co-expression of the strep-tagged viral protein and the TAP-tagged host proteins predicted to interact with it, followed by MS and western blot validation. Using this approach, all 18 HIV-1 proteins were shown with high confidence to be involved in 497 PPIs together with 435 host proteins [36]. A mixture of approaches, including tag-based AP and co-IP followed by MS identification, has
been used to identify the host proteins that interact with the HIV pre-integration complex, a key nucleoprotein required for the insertion of the reverse-transcribed viral DNA [38]. MS-based experiments were performed using infected CD4+ human cells.

CONCLUSION

It has been described in this review that protein interaction networks can elucidate the molecular basis of diseases, which in turn can appraise methods for prevention, diagnosis, and treatment. When the properties of these protein networks have been analyzed, novel higher order structures have been revealed. Therefore, it can provide an opportunity to interpret complex biological behaviors and alterations (in network dynamics) associated with complex diseases such as cancer and autoimmune diseases. These network relationships suggest a novel means of developing molecular therapies where the network is the target of therapy rather than individual molecules within the network. Hence, we expect that such systemic vantage of view should be applicable to complex diseases such as cancers and autoimmune diseases which are needed new efficient diagnosis and therapies and offers new opportunities for enhancing our understanding of complex diseases.

REFERENCES

