ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
"Antonio Ruberti"
CONSIGLIO NAZIONALE DELLE RICERCHE

P. Bertolazzi, M.E. Bock, C. Guerra, P. Paci,
D. Santoni

ON THE INTEGRATION OF PROTEIN-PROTEIN INTERACTION NETWORKS WITH GENE EXPRESSION AND 3D STRUCTURAL DATA: WHAT CAN BE GAINED?

R. 7, 2013

Paola Bertolazzi – IASI-CNR, Roma, Italy.

Mary Ellen Bock – Department of Statistics, Purdue University, West-Lafayette, IN, USA.

Concettina Guerra – College of Computing, Georgia Institute of Technology, Atlanta, GA, USA.

Paola Paci – IASI-CNR, Roma, Italy.

Daniele Santoni – IASI-CNR, Roma, Italy.

ISSN: 1128–3378

Collana dei Rapporti dell'Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti", CNR

viale Manzoni 30, 00185 ROMA, Italy

tel. ++39-06-77161
fax ++39-06-7716461
email: `iasi@iasi.cnr.it`
URL: `http://www.iasi.cnr.it`

Abstract

The biological role of proteins has been analyzed from different perspectives, initially by considering proteins as isolated biological entities, then as cooperating entities that perform their function by interacting with other molecules. There are other dimensions that are important for the complete understanding of the biological processes: time and location. However a protein is rarely annotated with temporal and spatial information. Experimental Protein-Proteins Interaction (PPI) data are static; furthermore they generally do not include transient interactions which are a considerable fraction of the interactome of many organisms. One way to incorporate temporal and condition information is to use other sources of information, such as gene expression data and 3D structural data. Here we review work done to understand the insight that can be gained by enriching PPI data with gene expression and 3D structural data. In particular, we address the following questions: Can the dynamics of a single protein or of an interaction be accurately derived from these data? Can the assembly-disassembly of protein complexes be traced over time? What type of topological changes occur in a PPI network architecture over time?

1. Introduction

A protein-protein interaction network for an organism represents physical interactions among proteins. Interactions are obtained through a variety of experimental and computational techniques, including Yeast-2-Hybrid (Y2H) [24] and Co-ImmunoPrecipitation (CoIP). Details on the experimental technologies can be found in a recent survey [15]. PPI networks are static since they do not show changes in interaction status between two proteins over time. In the human proteome as well as in other organisms a large number of interactions are transient [12], they occur for a limited time and are reversible, i.e. the two proteins disassociate and are no longer in physical contact. The interface involved in a transient interaction may be used by multiple partners at different times. Transient interactions are typically promoted by a set of conditions, such as phosphorylation and more generally post-translational modifications, or require localization to specific areas of the cell or are induced in response to a signal. An example of a transient interaction is that of a kinase binding a substrate within a particular signaling pathway. Examples of permanent interactions are those among proteins participating in a stable network complex that perform some specific function, for instance the 20S Proteasome.

Experiments to obtain interaction data at different points in time and under different conditions are currently too costly. Thus integrative computational approaches have been designed to add temporal and condition information into the PPI networks. One class of approaches integrates gene expression data and PPI data [25], taking advantage of the large amount of gene expressions accumulated over the past years for different organisms and under different conditions and time points. The dynamics of interactions, in the absence of direct experimental information, are derived from gene co-expression as follows: if two proteins are physically interacting in a PPI network and the two proteins also are co-expressed during a certain time interval during the cell cycle, then they are assumed to be interacting during that particular time interval. However this assumption is not always valid since an indirect interaction among the co-expressed genes may exist due to the presence of one or more intermediaries [56].

A second class of computational approaches to discriminate between transient and permanent interactions exploit 3D structural information as provided by X-ray crystallography and NMR and available from the Protein Data Bank (PDB) [8, 40, 50]. For a hub protein, i.e. a protein with a large number of interactions partners, not all interactions can take place simultaneously due to the limited surface area of the protein. Structural data and computational methodologies for structural comparison allow one to determine, for a given protein, which interactions can occur at the same time and which are mutually exclusive. Mutually exclusive interactions are obviously transient, while the interactions that can occur simultaneously may be permanent interactions. This condition can be verified by injecting into the analysis other data, for instance gene expression data. This approach, although powerful, is limited by the number of available structures of proteins in complex with other proteins.

The dynamics of individual proteins, not just of their interactions, have been analyzed by integrating gene expression data with PPI networks. Proteins may not be active at all times and under all conditions. Thus, some methods have explicitly traced the gene expressions over discrete time intervals, leading to a classification of gene products into *static* and *dynamic* proteins. Static proteins are defined as those expressed during the entire interval, while dynamic proteins appear only at some specific times. This definition generally relies on a chosen criterion and threshold to determine when a protein is present.

Another popular but controversial classification of proteins based on gene expression data in connection with PPI networks is in [22] where the dynamic properties of a protein were derived from the correlation of its expression profile with respect to its interacting partners in the network. This led to a separation of hub nodes, i.e. nodes with high degree, into *party* and *date* hubs: party hubs are highly correlated in their mRNA expression with their partners while date hubs show lesser correlation. Party hubs are believed to have permanent interactions, while data hubs are believed to have transient interactions.

It is known that proteins perform their function not in isolation but in cooperation with other proteins. Thus it is of interest to look at the dynamic behavior of groups of interacting proteins, or protein complexes. Much attention has been devoted recently to the process of assembly and disassembly of proteins complexes over time and to the way in which the same protein participates in different complexes at different times. Work in this area includes also the use of temporal information to better predict clusters of proteins that likely correspond to functional complexes.

A mathematical structure that supports these studies is the dynamic network that combines PPI data and gene expression data. Mathematically, the PPI network is a graph with each protein as a node. A graph G(V, E) consists of a set of nodes V and a set E of pairs (u, v), $u, v \in V$, called edges. In a PPI network an edge connects two nodes if a physical interaction has been observed between the two proteins. A dynamic network is a set of subnetworks derived from the original PPI network, each subnetwork corresponding to a time interval and representing all proteins and protein interactions detected during that time interval. A time-series of gene expression data combined with PPI data generates this series of subnetworks. A node of a subgraph corresponds to a protein. An edge connects two proteins if they are co-expressed at that time or time interval in addition to being connected by an edge in the static PPI network. Recall that co-expression is assumed to indicate common activity and therefore actual interaction. An alternative, but equivalent, mathematical way of looking at a dynamic network is as a single graph where the edge connecting two proteins is labeled by the points in times when both proteins are expressed.

The differences among the subnetworks composing a dynamic network are used to analyze the dynamics of a protein with respect to its immediate neighbors, the changes of network functional modules or complexes as well as of the topological properties of the whole network. The same type of analysis can be conducted and applied to expression data collected in different experimental conditions, for instance healthy/unhealthy individuals, different tissues, different metabolic conditions.

The dynamics of PPI networks has been the subject of a number of surveys [30, 38, 44, 51]. In this paper we illustrate different points of view on the issue of dynamics in PPI networks. We observe that dynamics can be studied concentrating either on single proteins or on subnetworks of PPI networks that vary in dimension and topology. Moreover, the gene expression data are analyzed both considering the absolute values of single genes or the correlation between two genes of the transcriptome. Consequently, we review the literature following this schema. In section 2 we survey papers investigating the dynamics of single proteins and interactions. Specifically, in section 2.1 we discuss the use of time-series genes expression values to identify static and dynamic proteins and the use of correlation of expression values to classify hub proteins into date and party hubs; in section 2.2 we review computational approaches to identify transient and permanent interactions based on 3D structural data. In section 3 we report work on the dynamics of different subnetworks, either modules or complexes, using gene expression correlation data to find possible rewiring of the static PPI network. In section 4 we present some results showing how the dynamics affects the entire PPI network, leading to a modification of a number of global and local properties of the network.

2. Dynamics of single proteins and interactions

The dynamics of proteins and interactions can be analyzed from two different perspectives. First, not all proteins may be active in a given cell at a given time and under a specific condition. Gene expression data acquired at different points in time and under different conditions provide a temporal annotation of proteins. Second, a protein may use the same interface to bind multiple proteins; thus such bindings cannot occur at the same time. Structural data and computational techniques enable to discriminate among simultaneously possible and mutually exclusive interactions, providing insight into the dynamics of interactions.

2.1. Dynamics from PPI and gene-expression data

Static and Dynamic Proteins

Although there is not a unique definition of a dynamic protein, there are ways of identifying such proteins. For instance, using time series of gene expression data, the timing of the transcription of genes allows one to distinguish between dynamic proteins, which are periodically expressed, and static proteins, which are constitutively expressed [16]. Dynamic proteins, in other words, show a peak of expression only at certain points in time while static proteins do not have a clear maximum.

The notion of static/dynamic can also be derived from condition-dependent gene expression data. Large datasets of genes of Saccharomyces C. with condition-dependent expression were analyzed in [29]. The expression variance (EV) across all experiments conducted under different conditions was calculated for each gene; this lead to classification of gene products into static, those with low variance, and dynamic, those that have a relatively high variance. A comparison of EVs of proteins with the EVs of their neighbors revealed that proteins have similar expression dynamics as their immediate neighbors in the protein interaction network.

A dynamic circadian interactome, reflecting the time-of-day dependent organization of the network, was constructed in [55] by a combining experimentally derived protein-protein interactions with time-of-day-dependent expression data. In the circadian oscillator, many of the known PPIs also happen predominantly at specific times of the day, e.g. PER/CRY complexes bind to CLOCK/BMAL1 in the late night, demonstrating the need of precisely timed PPIs for the circadian clockwork. The authors systematically mapped PPIs among 46 circadian components using high-throughput Y2H and co-immunoprecipitation experiments in human cells, thus identifying some so far uncharacterized interactions.

Party and Date hub classification and its relations to other biological and topological properties

This classification was one of the first attempts at introducing some temporal and spatial characteristics into an otherwise static set of data by using gene co-expression data. Han et al. [22]. examined the extent to which hubs in the yeast interactome are co-expressed with their interaction partners. They defined hubs as proteins with degree at least 5, where "degree" refers to the number of links emanating from a node. Based on the averaged Pearson correlation coefficient (avPCC) of expression over all partners, they concluded that hubs fall into two distinct classes: those with a low avPCC (which they called date hubs) and those with a high avPCC (so-called party hubs). They inferred that these two types of hubs play different roles in the network modularity: party hubs are thought to coordinate single functions performed by a group of proteins that are all expressed at the same time, whereas date hubs are described as higher-level connectors between groups that perform varying functions and are active at different times or under different conditions.

Date and party hubs have different topological properties. It was observed that removing date hubs produces a huge effect on the whole network connectivity leading to a rapid drop in the characteristic path length while the removal of party hubs is the same as randomly removing nodes [22]. Moreover, it has been shown that date hubs participate in more genetic interactions and evolve more rapidly than party hubs [9].

This hub classification is questioned in the papers [2, 7] where the bimodal distribution of the hubs is considered an artifact due to the algorithm used to normalize microarray expression data or to the kernel function used to smooth the histogram of the avPCC distribution. In addition, Batada et al. showed that there is no evidence for a lower rate of evolution of party hubs and any weak tendency for party hubs to evolve slower was accounted for by their abundance. The authors found that differences in abundance explain all difference in rates of evolution between date and party hubs, concluding that evolutionary rate differences do not support the date/party distinction. In the same paper the authors questioned also the differential deletion effect of party and date on the network integrity. The authors randomly swapped date and party hubs of the same connectivity and found the same deletion profile as the original unswapped case. These finding corroborated their hypothesis that the differential deletion effect is not solely due to inter- versus intra-module positioning but is sensitive to removal of just a few extreme hubs. In [2] the party and date hub classification is compared with a topological counterpart in which the role of the proteins is assigned on the basis of network topology rather than on the basis of expression data, as formulated by Guimer‘a and Amaral [20]. In this view the classification of nodes in a modular network rely on intra- and inter-module connectivities. According to this latter classification, date hubs should correspond to connectors between different clusters while date should represent provincial node connecting only node within the same cluster. However Agarval et al. showed that the two classifications are not consistent in two dataset they analyzed [2].

However, the idea that network hubs may fall into discrete classes is supported also in [26]. The authors hypothesized that date hub proteins have one or two interaction interfaces, but many partner proteins while party hub have many interaction interfaces but only one partner protein per interface. Each interface of date hub must interact with its partners in a mutually exclusive manner, establishing a hierarchy of interactions. Changes in the abundance of date hub proteins may have great consequence for the interactome as opposed to changes in the abundance of party hub which are supposed to produce a lesser effect on the interactome. Wilkins et al. in [57] analyzed this issue in relation to changes in the protein-complex assembly and function. These effects are not identical for all proteins and will differ depending on the number and type of interaction partners that a protein has, corroborating the hypothesis of a dichotomy date/party of network hubs [57].

Another classification of hub proteins based on co-expression profiles but restricted to a specific tissue is in [46]; proteins of the human interactome fall into two categories: intermodular and intramodular proteins. Intermodular hub proteins are co-expressed with their interacting partners in a tissue-restricted manner and intramodular hub proteins that are co-expressed with their interacting partners in all or most tissues.

For an extensive discussion of the controversy on party and date hub classification see [10, 13].

2.2. Dynamics of interactions from 3D structural data

Much structural data of proteins have become available in recent years, obtained by experimental techniques and by homology modeling. These data have enabled the construction of the 3D structure of the whole interactome of several organisms [35]. When experimental data related to a specific interaction is not available, it can be inferred from the structures of homologous proteins [4, 3, 5]. In other words, the 3D structure of a given protein-protein interaction is used as a template to model all the interactions that involve homologous proteins and for which the binding has been experimentally confirmed. The advantages of enriching PPI networks with a protein structural dimension have been discussed in a recent survey [13]; they include the prediction of new protein-protein interactions as well as uncovering new interesting features of different node types, such as disorder and evolutionary features.

From a structural perspective, proteins are classified in single (or singlish)-interface and multiple-interface. Single-interface have only a few interaction interfaces (two at most) that may be used by multiple partners at different times since these interactions are mutually exclusive. By contrast, multi-interface hubs allow simultaneous interactions. Statistically significant differences exist for these two classes of proteins in terms of important biological features. Singlish-interface hubs tend to be enriched in signaling proteins, whereas multi-interface hubs are often present in protein complexes [26].

Temporal quality of an interaction is inferred from this classification by considering the interactions of a single interface protein as generally transient, whereas those of multi-interface proteins are believed to be enriched in permanent associations. In the latter case, gene expression data can help determining if two proteins binding at different interfaces are actually simultaneous.

The large and increasing number of protein three-dimensional structures in the PDB provide a means to predict interactions that are mutually exclusive. The underlying principle of some of the approaches discussed below is the following: if two proteins bind a common partner and their interfaces are structurally similar, then they cannot bind such partner simultaneously.

Based on this principle, a structural interaction network (SIN) of the yeast was derived in [26, 28] in which many edges are annotated with structural information. They are labeled as simultaneous possible or mutually exclusive on the basis of sequence homology of the connected nodes to existing structurally defined complexes and of the presence or not of distinct interfaces.

Other approaches have relied only on structural homology to identify mutually exclusive interactions among all the structures present in the PDB. In [40] structural comparison is applied to the identification of the specific regions on the surface of proteins involved in the recognition process. The authors have analyzed PPI networks and retrieved sub-networks in which at least three proteins of known structure interact with a central protein. The surface residues of such proteins are compared and their structural similarity is used to predict which residues are likely to be in the binding interface of the proteins. The procedure, applied to the PPI networks of several organisms, showed satisfactory accuracy in predicting mutually exclusive interactions, when evaluated against known cases. As for the coverage of such method, i.e. percentage of sub-networks containing a sufficient number of proteins of known structure in the PPI maps, it varies substantially for different organisms, but it is generally above 30% and will certainly increase as more structures become available.

The usefulness of structural information to add the time dimensionality to a network is illustrated in [50] for the case of the hub protein p53, a protein with hundreds of links implicated in multiple signalling pathways in connection with human cancer. A structural comparison algorithm applied to a large number of available interfaces in the PDB predicts the interfaces of p53 with its interacting partners. This in turn reveals which interactions can co-exist and which are mutually exclusive.

A structural feature that has often been associated with transient interactions is intrinsic disorder. Disordered regions, i.e. regions without a rigid three-dimensional structure, provide the flexibility required by highly reversible yet highly specific interactions [54]. The tendency of a protein to bind to many other proteins requires some versatility in its conformation, as a number systematic studies of the available topological and predicted disorder data have confirmed. The analysis on single and multi-interface hubs revealed that single interface hubs are enriched for disorder while multi-interface hubs are not [28]. It was found that disorder is more present in complex organisms: in the human proteome about one third of the proteins residues are in disordered regions as apposed to only a few percent in other less complex species [18]. This implies that in the human proteome the presence of dynamically fluctuating interaction patterns [47].

Another type of dynamic behavior is considered in [8] where an additional dimension is added into molecular networks: dynamic conformational changes. Alternate conformations of proteins structures in human and yeast were extracted from the PDB databank and mapped into the Dynamic Structural Interaction Network (DynSIN) to determine the dynamic conformational changes occurring at their interfaces with different partners. The network allows one to derive a number of properties of the interactions, for example that transient associations involve smaller conformational changes than permanent ones.

3. Dynamics of protein complexes

Recently, the attention has shifted towards the analysis of the dynamics of groups of proteins (complexes, modules, or pathways) rather than of individual proteins or protein-protein interactions. It is known that some complexes assemble at a specific time and location to perform a certain function and then disassemble, while other complexes are stable. Furthermore, some proteins with transient associations are part of several complexes at different stages of a cellular process. However, the temporal and spatial information is rarely available in existing databases where protein complexes are essentially static entities.

The difference in the dynamic behavior of groups of interacting proteins was first outlined by Spirin and Mirny [42] who introduced the concept of functional modules vs protein complexes. The proteins in a complex interact with each other at the same time, by contrast, the proteins of a functional module bind to each other at a different times to carry out a specific cellular process.

To obtain insight into functional modules and complexes an integrative analysis of gene expression and interaction data has been performed by various authors in the last decade. Three main questions have been addressed: 1) What is the composition of complexes in terms of static/dynamic proteins? 2) Is there a relationship between the similarity in genes co-expression profiles and complex membership? In other words, do proteins in a complex have a significantly higher level of co-expression than random proteins? 3) Does the integration of physical and gene information lead to an accurate prediction of protein complexes through the study of their dynamics?

Characterization of complexes in terms of static/dynamic proteins

One of the early approaches to the analysis of dynamics of protein complexes is reported in [16]. They traced the formation of several complexes during the yeast cell cycle by combining reliable PPI data with information on the timing of transcription of several genes. They observed that most complexes consist both of dynamic proteins as well as static proteins. Furthermore, their analysis enabled the annotation of some uncharacterized proteins and modules. In [29] the positioning of proteins within modules was analyzed ; it was observed that large modules are either static, i.e. composed of static proteins and always present in the cell, or dynamic, i.e. composed of dynamic co-regulated proteins and expressed in a condition dependent manner.

The authors of [43] study the process of assembly-disassembly of proteins in a time of reference corresponding to the four phases of the cell cycle. It is observed that a subset of proteins can participate in a complex in one cycle and then in another at a different cycle. Constant gene co-expression of a group of proteins across all cycles is interpreted as re-usability, i.e. participation in different complexes at different points in time. While the majority of interactions connect static proteins, links between a static and a dynamic protein are also present indicating a transient interaction. It is this feature that is interpreted as re-usability or multi-purpose protein.

Gene co-expressions in complexes

That co-expression implies functional relationships was already observed in [19]. The general question of whether there exists a relation between the similarity in gene co-expression profiles and complex membership was addressed in [25] focusing on a set of known protein complexes of S.Cerevisiae. Based on correlated patterns of gene expression over a time course and across different conditions, a classification of complexes in *permanent* and *transient* was derived for large MIPS complexes. Significant co-expression of genes was observed in many permanent complexes, and on average much smaller correlation strength was detected in transient complexes. This result was confirmed in [48].

In [32] this relationship was investigated for the yeast network taking into account the size of the complexes. It was found that large complexes have a high level of gene co-expression. On the contrary, the average degree of co-expression in small complexes is not much higher than that from a pair of randomly selected genes. Genes of large protein complexes, such as the mitochondrial ribosomal proteins and the cytoplasmic ribosomal proteins, with high within-complex co-expression, have a strong influence on other genes since the vast majority of them are affected by the coherent expression of these large protein complexes. A similar result relating the size of complexes to the similarity in co-expression profiles was reported in [39] for the Homo Sapiens. They found that the gene expression of protein subunits is more likely to be similar for large complexes and that this is statistically significant.

Using gene expression data to identify protein complexes

Numerous computational approaches exist for the identification of clusters of proteins in static PPI networks that likely reflect the functional classification of proteins. Similarly, clustering algorithms have been applied to genetic networks, to identify groups of genes with high values of co-expression likely related to functional similarity. Only recently, however, clustering methods have been designed to integrate different types of data, from gene expression to interaction data, for the accurate identification of complexes and their dynamics.

Genetic, metabolic, regulatory and interaction information was analyzed in [48] to find functional modules. Basically, a cluster identified in one particular network was checked to see if its components were significantly related in another network. For example, having identified a module of the protein interaction network or using a protein complex they calculated the significance of the correlation strength of the corresponding gene expression profiles. In the case that a group or subset of the data correlates in a dense co-expression subnetwork, unknown genes that are members of such a subnetwork are candidates for interaction. Zhang et al [59] integrated various biological interaction types, such as protein-protein interaction, genetic interaction, transcriptional regulation, sequence homology, and expression correlation. They identified in the database MIPS pairs of known complexes with many inter-complex genetic interactions that constitute higher-order network structures or *network themes* tied to specific biological phenomena. In [41] co-expression data and PPI networks are integrated and a cluster procedure is applied separately to both; the clusters are then merged in a special graph and cliques of this graph are identified. The cliques are shown to be significantly enriched in known pathways.

Some integrative approaches [31, 45, 56] rely on the explicit construction of a dynamic network from gene expression and PPI data, i.e. a series a subnetworks each corresponding to a specific time point or time interval. In a subnetwork all the interactions are activated at the same time. A crucial aspect of the dynamic network construction is the choice of a threshold to determine the level of expression of genes products to be considered biologically significant, so that genes with values below the threshold are discarded. In [31, 45] the original static PPI network of S.cerevisiae retrieved from DIP is split into a series of time-sequenced subnetworks (TSNs) corresponding to different phases of the yeast metabolic cycle. The gene-expressing profiles of yeast are obtained from [49], which contains approximately 7000 genes under 36 different time points. A large fraction of these gene products are present in the PPI network. A fixed threshold is used to filter gene products at each time point. Only the transcripts whose expression levels are greater than a fixed threshold (set to 0.7) remain. With this choice of threshold, about 40% raw transcripts with low expressing levels are removed. In each of the 36 subnetworks, links represent interactions that are activated at the same time. Then a cluster procedure identifies within each subnetwork protein complexes which are then traced in different networks to study their dynamic behavior. Those complexes that are found in a number of subnetworks correspond, according to the classification of [42], to protein complexes. The results of their procedure provides a better match to protein complexes from MIPS than the results of a similar clustering procedure applied to a static PPI network.

In [56] a different approach is taken to the construction of a dynamic network, specifically in the selection of the threshold to filter genes. Active time points of each protein in a cellular cycle are identified, using an active threshold for each gene according to the characteristics of its expression curve. This is based on the observation that different proteins in a single cell have different expression patterns. Furthermore, there exist proteins with important role whose expression levels are very low during the whole cycle. A fixed global threshold would miss such proteins. According to the the statistics in [56] 23-45% proteins are active at a time point and most proteins are active in about half of cellular cycle. The yeast PPI network is retrieved from DIP [52] and two gene expression sets from GEO [1] are used to derive active times for proteins. The resulting dynamic network can be represented by 12 subnetworks corresponding to 12 time points. In each subnetwork, the proteins are active at the time point, and the interactions that appear at the time point are also maintained. A clustering algorithm is then applied to the DPIN network and is evaluated against the algorithms MCL [58], CPM [37] that detect clusters in a static PPI networks. The performance of each algorithm on DPINs outperforms those on other networks in terms of matching with known complexes, sensitivity, specificity, f-measure, and accuracy.

Related to these studies although not specifically focusing on interaction dynamics are the approaches to functional modules identification that integrate physical interaction data with genetic information [11, 14, 34].

4. Dynamics of whole networks

The dynamics of biological networks in terms of global topological changes as well as in the occurrence of specific motifs has been largely overlooked due to the lack of complete and reliable data representing biological entities and their relations over time. One exception is the statistical analysis of the dynamics of transcriptional regulatory networks. For instance, in [33] the transcriptional network of Saccharomyces cerevisiae was integrated with gene-expression data for multiple conditions. Both global and local topological properties are related to specific functional properties of transcription factors and their dynamics under 5 different conditions. Transcription factors generally are hub nodes of the networks and a few of them are shown to have permanent interactions while the others have transient interactions. These latter factors rewire the network, altering their interactions to varying degrees.

As for PPI networks, the analysis of the changes of global and local topological properties was performed in [23] for the networks of E.coli obtained at different times under different environmental conditions. The networks were obtained at 20 minute time intervals after exposure of wild-type and SOS E.coli to UV. Although no significant changes could be observed globally (for instance the diameter of the networks is identical and the average degree very similar), some small perturbations were detected that might yield a useful insight into the cellular response. Interestingly, the shortest path length between a few pairs of nodes varied by more than 3, a value considered significant in the paper, although the majority of shortest path lengths stayed the same. As expected, a number of genes and their protein products are expressed only in some of the networks in agreement with their known biological role. On the other hand, a large fraction of hubs (defined as nodes with degree higher than 60) were present in all conditions, supporting the claim that hub nodes are more essential than non-hub nodes. The comparative analysis of the four networks revealed some changes that are physiologically relevant and that could not be revealed by the analysis of gene expression alone.

5. Conclusions

Interaction data accumulated through a variety of experiments from small scale to high-throughput have provided much insight into the cellular mechanisms, allowing to look at proteins in terms of the biological processes in which they are involved rather than as isolated entities. Networks with varying degrees of accuracy and completeness are now available for various organisms, including yeast, human, drosophila. Much information is however missing from these graphical representations. For instance, transient interactions are often not included because of the difficulty of acquiring these data. More in general, proteins and interactions are rarely annotated with temporal and condition information. Hopefully, the limitations in the acquisition of direct interaction data over time will be overcome in the near future.

In the meantime, to cope with the current lack of experimental data capturing the dynamics of the interactions, other sources of information have been added to the PPI networks. Gene expression data and structural 3D data are a natural complement to PPI maps. Large amounts of gene expression data are already available under different conditions and at specific times or time intervals and their rate of increase is progressing. Here we have discussed the use of such data in determining when a protein is active, when its interactions with other molecules occur and its participation in the assembly/disassembly of complexes or modules over time. Structural data have also increased significantly over the past decades and are expected to grow although not at the same pace as genetic data. Among other things, they provide the basis to discriminate between mutually exclusive and simultaneously possible interactions. Overall, much gain is achieved from the integration of all these data in terms of knowledge of biological processes and more is expected to be achieved.

Acknowledgements

P.B., P.P. and D.S. were partially supported by the FLAGSHIP InterOmics project (PB.P05) and by Sysbio project both funded by the Italian MIUR and CNR institutions.

References

[1] http://www.ncbi.nlm.nih.gov/geo/

[2] Agarwal S., Deane C.M., Porter M.A., Jones N. S. (2010) Revisiting Date and Party Hubs: Novel Approaches to Role Assignment in Protein Interaction Networks, PLoS Comput. Biol., 6, 6.

[3] Aloy P., Russell R.B., (2002) Interrogating protein interaction networks through structural biology. Proc. Natl. Acad. Sci. USA, 99, 5896-5901.

[4] Aloy P., Russell R.B. (2002) The third dimension for protein interactions and complexes. Trends Biochem. Sci. 27, 633-638.

[5] Aloy P., Russell R.B. (2006) Structural systems biology: modelling protein interactions. Nature Rev. Mol. Cell Biol., 7, 188-197, doi: 10.1038/nrm1859.

[6] Bandyopadhyay S, Kelley R, Krogan NJ, Ideker T (2008) Functional Maps of Protein Complexes from Quantitative Genetic Interaction Data. PLoS Comput. Biol., 4, 4, e1000065. doi:10.1371/journal.pcbi.1000065

[7] Batada N.N., Hurst L.D., Tyers M. (2006) Evolutionary and physiological importance of hub proteins. PLoS Comput. Biol., 2, 7, e88.

[8] Bhardwaj N., Abyzov A.,Clarke D., Shou C., Gerstein M.B. (2011) Integration of protein motions with molecular networks reveals different mechanisms for permanent and transient interactions. Protein Science, 20, 10, 1745-1754.

[9] Bertin l., (2007) Confirmation of Organized Modularity in the Yeast Interactome. Plos Comp. Biology, 5, 6.

[10] Bertolazzi P., Bock M.E., Guerra C. (2013) On the functional and structural characterization of hubs in protein-protein interaction networks. Biotechnol Adv., 31, 2, pp. 274-86, 2013.

[11] Beyer A., Bandyopadhyay S., Ideker T. (2007) Integrating physical and genetic maps: from genomes to interaction networks. Nature Rev. Genet., 8, 699-710.

[12] Brown K.R., Jurisica J. (2007) Unequal evolutionary conservation of human protein interactions in interologous networks. Genome Biology, 8, R95.

[13] Clarke D., Bhardwaj N., Gerstein M.B. (2012). Novel insights through the integration of structural and functional genomics data with protein networks. J. Struct. Biol., 179, 320-326.

[14] Collins S.R., Miller K.M., Maas N.L., Roguev A., Fillingham J, et al. (2007) Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature, 446, 806-810.

[15] De Las Rivas J., Fontanillo C. (2010) Protein-Protein interactions essentials: key concepts to building and analyzing interactome networks. PLoS Comput. Biol., 6, 6, e1000807.

[16] De Lichtenberg U., Jensen L.J., Brunak S., Bork P. (2005) Dynamic complex formation during the yeast cell cycle. Science, 30, 724-727.

[17] Du N., Zhang Y,. Li K., et al. (2012) Evolutionary Analysis of Functional Modules in Dynamic PPI Networks. ACM-BCB12, 250-257.

[18] Dunker A.K., Oldfield C.J., Meng J., Romero P., Yang J.Y., Chen J.W., Vacic V., Obradovic Z., Uversky V.N. (2008) The unfoldomics decade: an update on intrinsically disordered proteins. BMC Genomics, 16, 9.

[19] Eisen M., Spellman P., Brown P., Botstein D. (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95: 14863-14868.

[20] Guimerà R., Sales-Pardo M., Amaral LAN (2006) Classes of complex networks defined by role-to-role connectivity profiles. Nat. Phys., 3, 63-69.

[21] Gursoy A., Keskin O., Nussinov R. (2008) Topological properties of protein interaction networks from a structural perspective. Biochem. Soc. Trans., 36, 6, 1398-1318,

[22] Han J.-D.J., Bertin N., Hao T., Goldberg D.S., Berriz G.F., Zhang L.V., Dupuy D., Walhout A.J.M., Cusick M.E., Roth F.P., Vidal M. (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature, 430, 88-93.

[23] Hegde S.R., Manimaran P., Mande S.C. (2008) Dynamic Changes in Protein Functional Linkage Networks Revealed by Integration with Gene Expression Data. PLoS Comput. Biol., 4, 11, e1000237. doi:10.1371/journal.pcbi.1000237

[24] Ito, T. et al., 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. USA 98, 45694574.

[25] Jansen R., Greenbaum D., Gerstein M., (2002) Relating whole genome expression data with protein-protein interactions. Genome Res., 12, 37-46.

[26] Kim P.M. et al. (2006) Relating three-dimensional structures to protein networks provides evolutionary insights. Science, 314, 1938-1941.

[27] Kim W.K., Henschel A., Winter C., Schroeder M. (2006) The many faces of protein-protein interactions: a compendium of interface geometry. PLOS Comp. Biol., 2, 9, e124.

[28] Kim P.M., Sboner A., Xia Y., Gerstein M. (2008) The role of disorder in interaction networks: a structural analysis. Molecular Systems Biology, 4, 179.

[29] Komurov K., White R. (2007) Revealing static and dynamic modular architecture of the eukaryotic protein interaction network. Mol Syst Biol., 3, 110.

[30] Levy E.D. , Pereira-Leal J.B. (2008) Evolution and dynamics of protein interactions and networks Current Opinion in Structural Biology, 18, 1-9

[31] Li M., Wu X., Wang J., Pan Y. (2012) Towards the identification of protein complexes and functional modules by integrating PPI network and gene expression data BMC Bioinformatics, 13, 109, doi:10.1186/1471-2105-13-109

[32] Liu C.T., Yuan S., Li K.C. (2009) Patterns of co-expression for protein complexes by size in Saccharomyces cerevisiae. Nucleic Acids Research, 37, 2, 526532, doi:10.1093/nar/gkn972

[33] Luscombe N.M., Babu M.M., Haiyuan Y.H., Snyder M., Teichmann S., Gerstein M.(2004) Genomic analysis of regulatory network dynamics reveals large topological changes, Nature, 431, 308-312, doi:10.1038/nature02782

[34] Maraziotis I.A., Dimitrakopoulou K., Bezerianos A. (2007) Growing functional modules from a seed protein via integration of protein interaction and gene expression data. BMC Bioinf., 8, 408.

[35] Mosca R., Col A., Aloy P. (2012) Interactome3D: adding structural details to protein networks. Nature Methods, DOI: 10.1038/NMETH.2289

[36] Obayashi T., Hayashi S., Saeki M., Ohta H., Kinoshita K. (2009) ATTED-II provides coexpressed gene networks for Arabidopsis Nucleic Acids Res. 37 (Database issue), D987D991. doi: 10.1093/nar/gkn807

[37] Palla G., Dernyi I., Farkas I., Vicsek T., (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature, 435, 814-818.

[38] Przytycka T.M., Singh M., Slonim D.K (2010) Toward the dynamic interactome: it's about time. Briefings in Bioinformatics, 11, 1, 15. doi:10.1093/bib/bbp057.

[39] Kikugawa S., Nishikata K., Murakami K., Sato Y., Suzuki M., Altaf-Ul-Amin M.D., Kanaya S., Imanishi T. (2012) PCDq: human protein complex database with quality index which summarizes different levels of evidences of protein complexes predicted from H-Invitational protein-protein interactions integrative dataset. BMC Systems Biology, 6 (Suppl 2):S7 doi:10.1186/1752-0509-6-S2-S7

[40] Sanchez Claros C. Tramontano A. (2012) Detecting mutually exclusive interactions in protein-protein interaction maps. PLoS ONE, 7, e38765.

[41] Santon D.i, Swiercz A., mieko A., Kasprzak M., Blazewicz M., Bertolazzi P., Felici G., (2013) An integrated approach to combine expression data and protein-protein interaction networks and its application on Arabidopsis thaliana. (manuscript)

[42] Spirin V., Mirny L.A. (2003) Protein complexes and functional modules in molecular networks. Proc Natl Acad Sci, 100, 21, 12123-12128.

[43] Srihari S., Leong H.W. (2012) Temporal dynamics of protein complexes in PPI Networks: a case study using yeast cell cycle dynamics, BMC Bioinformatics, 13 (suppl 17), S16.

[44] Stelzl U., Wanker E. (2006) The value of high quality proteinprotein interaction networks for systems biology. Current Opinion in Chemical Biology, 10, 551-558.

[45] Tang X., Wang J., Liu B., et al. (2011) A comparison of the functional modules identified from time course and static PPI network data. BMC Bioinformatics, 12, 339.

[46] Taylor I.W., Linding R., Warde-Farley D., Liu Y. Pesquita C., et al. (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nature Biotechnology, 27, 199-204.

[47] Tompa P. (2012) On the supertertiary structure of proteins. Nat Chem Biol, 8, 597-600.

[48] Tornow S., Mewes H.W. (2003) Functional modules by relating protein interaction networks and gene expression. Nucleic Acids Res, 31, 6283-6289.

[49] Tu B.P., Kudlicki A., Rowicka M., McKnight S.L. (2005) Logic of the yeastmetabolic cycle: temporal compartmentalization of cellular processes. Science, 310, 1152-1158.

[50] Tuncbag N., Kar G., Gursoy A., Keskin O, Nussinov R. (2009) Toward inferring time dimensionality in protein-protein networks by integrating structures: the p53 example. Mol Biosyst., 5. 12, 1770-1775.

[51] Vidal M., Cusick M.L., Barabasi A.L. (2011) Interactome Networks and Human Disease. Cell, 144, 986-988.

[52] Xenarios I., Salwnski L., Duan X.J., Higney P., Kim S.M., Eisenberg D. (2002) DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. Nucl. Acids Res., 30, 1, 303-305.

[53] Yu H., Braun P., Yildirim M.A., Lemmens I., Venkatesan K., et al. (2008) High-quality binary protein interaction map of the yeast interactome network. Science. 322, 104-110.

[54] Yura K., Hayward S. (2009) The interwinding nature of protein-protein interfaces and its implication for protein complex formation. Bioinformatics, 25, 3108-3113.

[55] Wallach T., Schellenberg K., Maier B., Kalathur R.K.R., Porras P, et al. (2013) Dynamic Circadian ProteinProtein Interaction Networks Predict Temporal Organization of Cellular Functions. PLoS Genet. 9, 3, e1003398. doi:10.1371/journal.pgen.1003398

[56] Wang J., Peng X., Li M., Pan Y., (2013) Construction and application of dynamic protein interaction network based on time course gene expression data. Proteomics, 13, 301-312, DOI 10.1002/pmic.201200277 301

[57] Wilkins M.R., Kummerfeld S.K. (2008) Sticking together? Falling apart? Exploring the dynamics of the interactome. Trends in Biochemical Sciences, 33, 195-200.

[58] Vlasblom J., Wodak S. J. (2009) Markov clustering versus affinity propagation for the partitioning of protein interaction graphs. BMC Bioinform., 10, 99-112.

[59] Zhang L.V., King O.D., Wong S.L., Goldberg D.S., Tong A.H., et al. (2005) Motifs, themes and thematic maps of an integrated Saccharomyces cerevisiae interaction network. J Biol., 4, 6.

16.

[60] Zotenko E., Guimares K.S., Jothi R., Przytycka M. (2006) Decomposition of overlapping protein complexes: A graph theoretical method for analyzing static and dynamic protein associations. Algorithms for Molecular Biology, 1, 7, doi:10.1186/1748-7188-1-7.