

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Scalable, Integrative Analysis and Visualization of Protein Interactions

David Otasek¹, Chiara Pastrello^{2,1} and Igor Jurisica^{1,3}

¹*Ontario Cancer Institute the Campbell Family Institute for Cancer Research,
University Health Network, Toronto, Ontario,*

²*CRO Aviano, National Cancer Institute, Aviano,*

³*Department of Computer Science and Medical Biophysics,
University of Toronto, Toronto,*

^{1,3}*Canada*

²*Italy*

1. Introduction

Biology offers a diversity of problems, leading to many computational biology workflows, including tasks where network visualization is helpful to interpret and analyse data. High-throughput screening techniques generate large amounts of data useful for the comprehension of the biological mechanisms underlying different diseases. The need for agile tools to handle such data and analyse it correctly has become continuously more evident.

Individual network visualization systems differ greatly in terms of the features and standards they support, and consequently the analyses they enable. Importantly, users have a broad range of skills and expectations, ranging from biology to computational biology. As a result, network visualization tools must satisfy diverse requirements and thus offer different user interfaces and features. In this role, they are also fundamental in helping scientists in different fields integrate their knowledge and their data in an interdisciplinary approach to research.

The number of '-omics' disciplines that use high-throughput techniques and that can benefit from a network approach are increasing. The diverse data that can be represented as a graph includes physical protein-protein interactions (PPIs), metabolic networks (Swainston et al., 2011), genetic co-expression (Helaers et al., 2011), gene regulatory networks (Longabaugh, 2012), microRNA-target (Shirdel et al., 2011) and drug-target associations (Morrow et al., 2010). In this chapter we focus on physical PPIs.

Proteins are key players in virtually all biological events that take place within and between cells and often accomplish their function as part of large molecular machines, whose action is coordinated through intricate regulatory networks of transient PPIs. The understanding of the interrelationships between molecules is the basis for an understanding of the behaviour of biological systems (Stein et al., 2011).

The analysis of the full proteome is possible with techniques such as mass spectrometry and protein microarray, which can be integrated with targeted approaches such as yeast-2-hybrid screen, immune precipitation and affinity purification. So far, PPI discovery methods are not accurate enough to be used alone, but the combination of different techniques can help to build an accurate interactome map (Remmerie et al. 2011). Still, this kind of analysis can only indicate that two proteins interact but does not reveal the molecular details or the mechanism of binding captured in high resolution three-dimensional (3D) structures, in which individual residue contacts are resolved and the interaction interfaces characterized. Moreover, they do not capture transient interactions and post translational modifications (PTMs) that can be addressed by techniques such as immobilized metal affinity chromatography (IMAC) mass spectrometry for protein phosphorylation analysis.

It becomes evident that the analysis of protein interactions is already a huge field with a plethora of data coming from different sources that can be improved by computational techniques and integrative network visualization and analysis. It is even more interesting to integrate PPI data with protein-target interaction data to have a wider view of the environmental context that influences network operations.

In this context, a pathway-centric analysis can help to elucidate the role and the importance of proteins in the context of the cell environment, specifically when the pathways can be related to the process/disease being studied. However, it is mandatory to be aware of the limits of this analysis, due to the cross-talk among pathways: a singular protein, in fact, can be associated or interact with multiple pathways so none of the pathways can be considered a single actor but rather a piece of a bigger puzzle (Kreeger & Lauffenburger, 2010).

Another intriguing aspect that protein-target interactions can describe is the relationship between protein exogenous molecules like drugs or toxins (Yu, 2011). The analysis of networks generated from drug-target and protein-target interactions can highlight different molecules that can be responsible of the response or resistance to a certain drug as well as alternative drugs that can target disease specific proteins.

2. Network visualization tools

There are dozens of applications available for the visualization of biological networks, each with its own focus, work-flow and tools (Pavlopoulos et al., 2008; Gehlenborg et al., 2010). We will describe some of the most common features and workflows involved in using these applications, with brief discussion of NAViGaTOR (Brown et al., 2009; McGuffin and Jurisica, 2009; Djebbari et al., 2011), Cytoscape (Smoot et al., 2011), VANTED (Björn et al., 2006) and VisANT (Mellor et al., 2004), four popular multi-platform biological network visualization applications.

2.1 Biological networks as annotated graphs

The most basic mathematical structure common to all of these applications is the graph, a collection of objects connected by links, referred to as nodes and edges. These objects are abstractions of real-world biological entities, where nodes could represent proteins, genes, molecules, drugs, etc. and edges could represent physical protein-protein, metabolic, or genetic interactions, microRNA to target associations, correlation, similarity relationships,

etc. Edges can be directed or undirected, weighted or not. In a case like gene regulation, Gene A may regulate Gene B, but the relationship may not be symmetric, meaning Gene B does not regulate Gene A. These models of biological networks have differing levels of support across various applications. An application may only support a small subset of node and edge types in order to specialize on one particular model, such as VisANT, which integrates many specialized tools for tasks such as Gene Ontology (GO) annotation, name resolution and online searches. Other applications may be more open-ended to provide support for as many models as possible, such as NAViGaTOR, Cytoscape and VANTED. The advantage of such a model is versatility, but it comes at the cost of having to manually define the nature of each node or edge via annotation.

To populate a graph within an application, the application must support one or more input formats. Often, the most basic level of input is either plain text or spreadsheet files such as Excel XLS format. For more graph-specific data, such as layout, GML can be used. To support more complex and structured biological data, several community standards exist: PSI-MI, BioPAX, and SBML.

Adding new nodes and edges to an existing graph can generally be done manually or by adding additional interactions from a supported database or file format. Some applications may have a workspace that supports concurrent, multiple graphs, which can then be combined or compared in various ways. Cytoscape and NAViGaTOR both support this type of workspace.

Once the graph is loaded within an application, a researcher may wish to add additional annotations, such as gene or protein expression, experimental confidence measures or Gene Ontology (The Gene Ontology Consortium, 2000) to their graph objects. Data from in-house sources must generally conform to the application used; generally, this is in the form of spreadsheets or text data with varying degrees of format flexibility. The researcher can also call upon more specialized data from public databases, such as UniProt, Entrez, KEGG or Genbank, either through the import of files or from direct access to the database through the application or a plug-in.

The amount of biological networks available to the researcher is ever expanding, and the size of the networks involved in many types of analysis is in order of thousands of nodes and edges. For example, the yeast interactome comprises 23,918 interactions according to DIP and 152,877 known and predicted interactions in I2D, the Interologous Interaction Database (<http://ophid.utoronto.ca/i2d>), an integrated database of PPIs from curated databases, experimental sources and predicted interactions (Niu et al., 2010; Brown and Jurisica, 2007; Brown and Jurisica 2005). While the researcher may only be interested in a small portion of the network in question, the scalability of an individual application and its analysis methods to networks of such size can be a considerable advantage.

2.2 Network visualization

Part of the challenge of visualizing a network is the laying out of the graph in a comprehensible manner. For smaller graphs, manual editing of node positions may be sufficient. With the aforementioned instances of graphs in the order of thousands of interactions, more robust tools are available with which to lay out a graph. Automated

graph layout algorithms, such as the force-directed and hierarchical, make the process easier, but often produce messy, uninterpretable graphs. Manual control over the placement of nodes and specialized tools for doing so are often necessary, from simple movement of single nodes to alignments in circles and lines to manipulate groups of nodes.

Algorithms for graph analysis are generally included in each application. Here, the number and type of analyses available are wildly variable. Algorithms can be used to find important graph properties, such as node degree, centrality, shortest paths, cliques and clusters. In addition, diverse biology-specific algorithms exist such as GeneMANIA (Montejo et al., 2010). Some applications may be designed specifically for one type of analysis while others contain a variety of analysis methods and in some cases allow for the addition of third party methods through plug-ins (NAViGaTOR, Cytoscape) or scripting languages (VisANT).

How an application chooses to visualize a graph is also variable. Nodes can be represented as anything from basic geometric shapes with variable size, color and transparency to application specific or user supplied bit-mapped images (Cytoscape, VANTED) or even other data visualizations such as bar charts (VANTED, VisANT). Edges can be straight, curved, displayed with various dot or dash schemes and can have variable widths, colors and transparencies. To make certain attributes readily visible, it is also possible in some instances to map an attribute to a visual property, such as color or size. All four of our example applications have different implementations of such mapping; the utility of a specific implementation is dependent upon the needs and competencies of the individual researcher.

Once the graph satisfies the requirements envisioned by the researcher, its state must be stored or exported. Proprietary formats are generally the norm for most programs, as visualization and data are often application specific and must be stored for later editing. Export formats often take the form of community standards (PSI-MI, BioPAX) and graphical exports. Graphical export is generally the final stage before publication. Usually, this can be done in bitmap (JPEG, TIFF, PNG, etc.) or vector (SVG, PDF, etc) formats, the latter being preferable for publication, as it can be resized and manipulated without loss of quality.

2.3 NAViGaTOR

NAViGaTOR (Network Analysis, Visualization and Graphing Toronto; <http://ophid.utoronto.ca/navigator>) is a network and graph visualization application with an emphasis on large graphs with integrated data (Brown et al., 2009). Data can be imported using diverse formats, ranging from community standards such as PSI-MI XML (Kerrien et al., 2007), BioPAX (Demir et al., 2010) or GML (Himsolt, 1996), to user-defined text files. Though the application is geared towards protein-protein interactions, the graph implementation within NAViGaTOR is not PPI specific, and can be used to model many types of real world or theoretical objects. Nodes and edges can have data associated with them, from simple numeric or text data to structured XML. Once imported, graphs can be combined from within a multi-graph workspace using combinations of cut, copy and paste operations. Additional data for the annotation of existing graphs can be imported using compatible files or online resources, such as I2D, cPath, or the one of the many online databases implementing the PSICQUIC web service.

Graphs generated by the above methods can quickly increase in size to thousands of nodes and edges. NAViGaTOR was designed with networks of this size in mind. While graphs this

size do create a demand for both memory and processing power to render, layout and navigate, the conservation of important paths and data is important to end-user analysis, particularly since most graphs of interest are subsets of a much larger interaction networks. NAViGaTOR approaches the problem of limited computing resources through the combination of a powerful OpenGL rendering engine through JOGL, and a suite of efficient layout, search and analysis tools. The JOGL rendering system gives the application access to the graphic processing power of the OpenGL compliant hardware of most graphics cards, allowing the application to use the CPU for more intensive graph operations.

NAViGaTOR supports several layout algorithms tailored for large graphs, including GRIP (Graph Drawing with Intelligent Placement) and several variants of the force directed algorithm. These algorithms come in both single and multi-threaded modes to take advantage of computers with multi-core CPUs.

When the structure and data contained within a graph are sufficient, the user can then interact with the graph, identifying significant nodes, edges or subsets of the graph using a variety of searches, spreadsheet tables and algorithms. Online or file supported databases can also be used to indicate known pathways and complexes within the data.

Users can highlight interesting structures within a graph with a variety of methods. Nodes and edges can be assigned visual properties to differentiate them from each other. Nodes can be given different colors, sizes, and highlighting styles. Edges can be given different colors, widths and styles and have the option to be rendered as user adjustable curves. Transparency can be used on both nodes and edges to either increase or decrease the visibility of graph objects.

The user can save the file in native NAViGaTOR format, GML, PSI-MI or delimited plain text. In addition, for presentation or publication purposes, the graph can be exported to one of several graphical formats, including JPEG, PNG, TIFF, SVG and PDF.

3. Iterative expansion of a protein interaction network

The increasing amount of data that can be collected from high-throughput analyses is accelerating research in the field of molecular biology; however, data of this type is also challenging due to its size. It can be used either for knowledge-based targeted analyses, meaning to improve the understanding of the role of an important well-known player in a specific field of interest (for example of BRCA1 in breast cancer), or unbiased analyses to understand the processes involved in a specific behaviour without a priori knowledge (for example, which genes/proteins are responsible for the poor survival of patients with pancreatic cancer?)

For our example, we have a list of potential interactors for a hypothetical protein of interest, PRO1, generated by computational PPI prediction. Also at our disposal are two meta analyses efforts, specifying the number of ovarian or prostate cancer related studies found in which the gene and its interactors were significantly deregulated. All other data will be collected from publicly available resources, including a PPI database, and a catalogue of drugs and their gene targets.

For our example, we will start with our experimental data in a tabular format. Data such as this can be obtained from any number of sources, from high-throughput experiments to

computational predictions. In our case, we have 21,302 predicted PPIs. Our analysis has produced a confidence metric associated with each interaction, ranging from 0 to 1.0. This confidence metric can be used to reduce the number of interactions we are dealing with to a more manageable size by removing lower confidence interactions. Our cut-off for high confidence will be 0.892, a value determined by cross validation. This leaves us with only 39 interactions, a far more manageable number for the next analysis steps. More complex filtering can be done through a simple spreadsheet application, such as Excel, or with a mathematical application such as R or Matlab.

At this point, we translate this data into a pair-wise table of PPIs, and import this table into NAViGaTOR. While NAViGaTOR supports several formats for loading interactions, we have chosen the tab-delimited format to facilitate easy translation from our original data. Other interaction data sets can be imported using community standard file formats, such as BioPAX, GML, PSI-MI XML and PSI-MITAB. Though these formats are harder to construct, they can contain more structured data, and facilitate easier data interchange among diverse programs and databases.

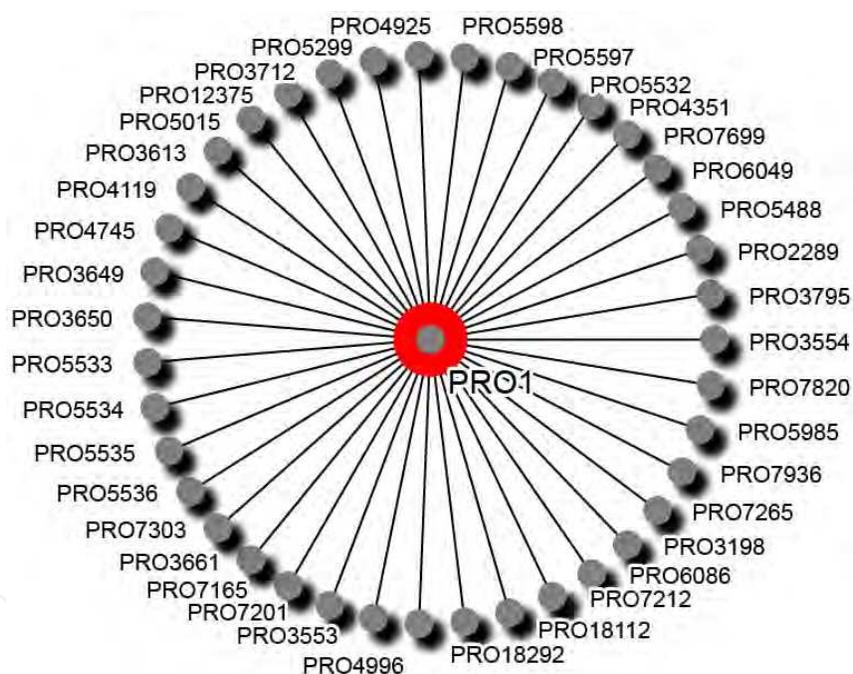


Fig. 1. Example graph containing hypothetical protein PRO1, with interactors loaded from experimental data. Tabular view of the data is available as a supplemental material (<http://www.cs.utoronto.ca/~juris/data/intech12/>).

Loading our pair-wise data, we get a very basic view (Figure 1). The visualization of this network at this stage is a spoke diagram with PRO1 in the center, and offers little information to the researcher that could not have been seen through a simple spreadsheet. We already have data regarding 39 interactions in the form of the confidence metric imported from our initial study. This can be mapped to one or more visual attributes using NAViGaTORs filter framework. In this case, we can make the highest confidence interactions more visible by applying a filter to map confidence to both edge width and transparency (Figure 2).

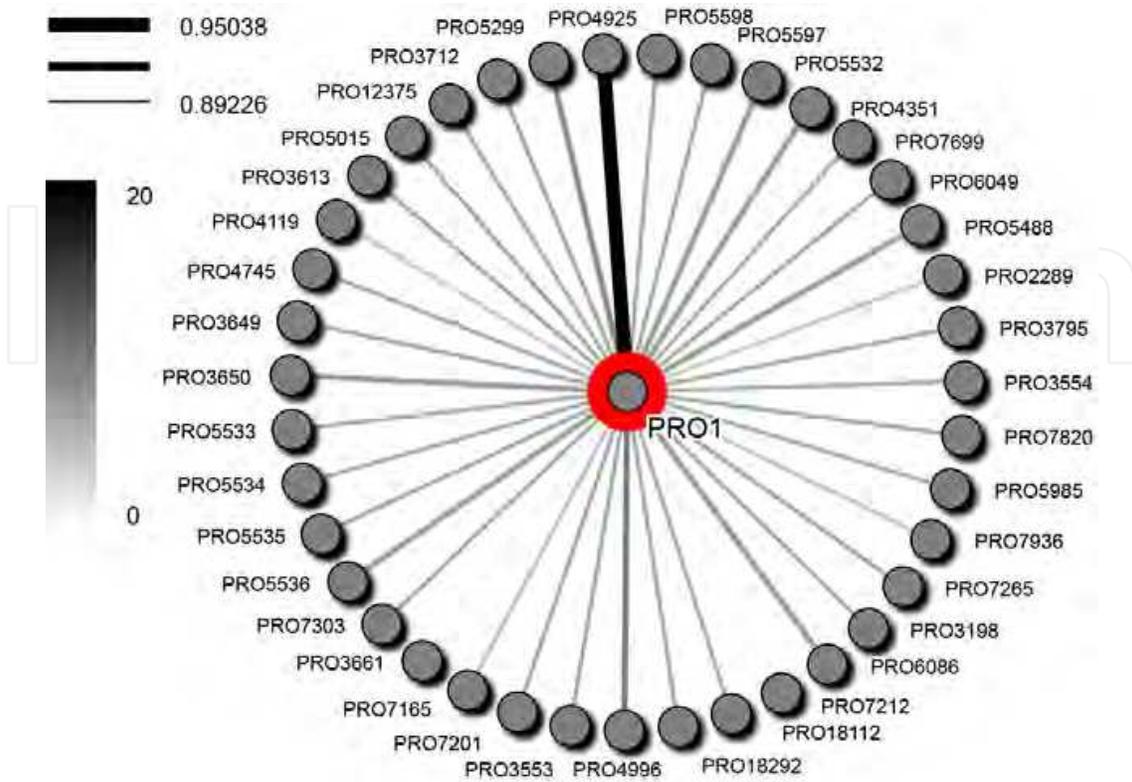


Fig. 2. Example graph with experimental interaction confidence mapped to edge width and transparency.

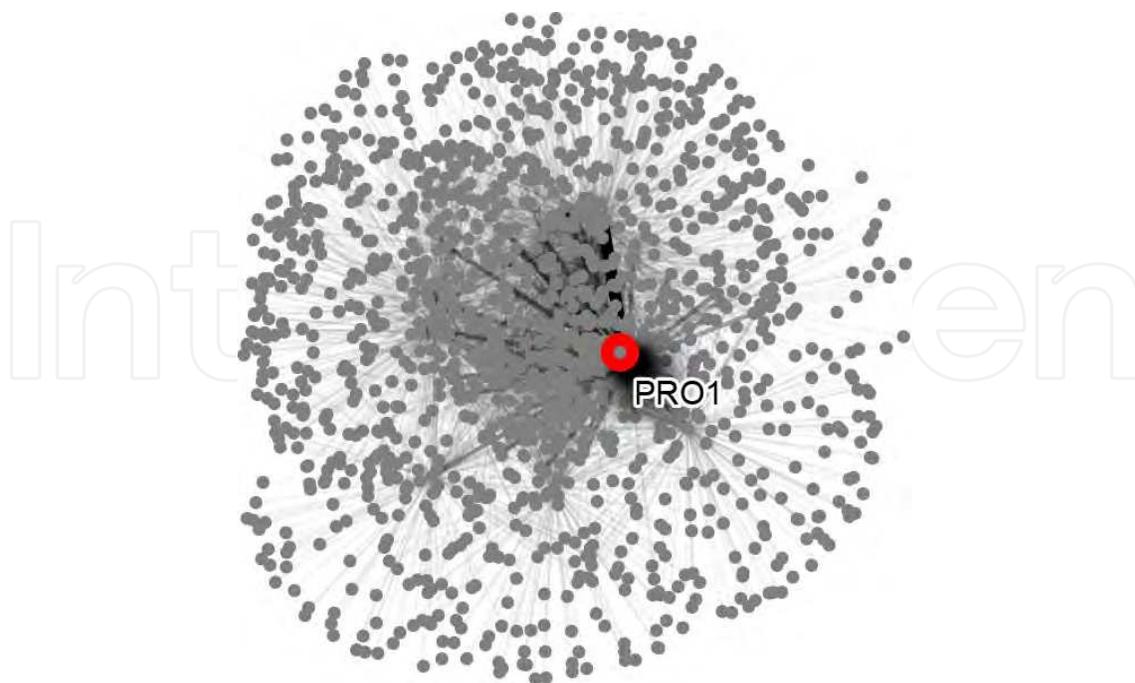


Fig. 3. Example graph enriched with interactions loaded from I2D, and laid out using the GRID algorithm.

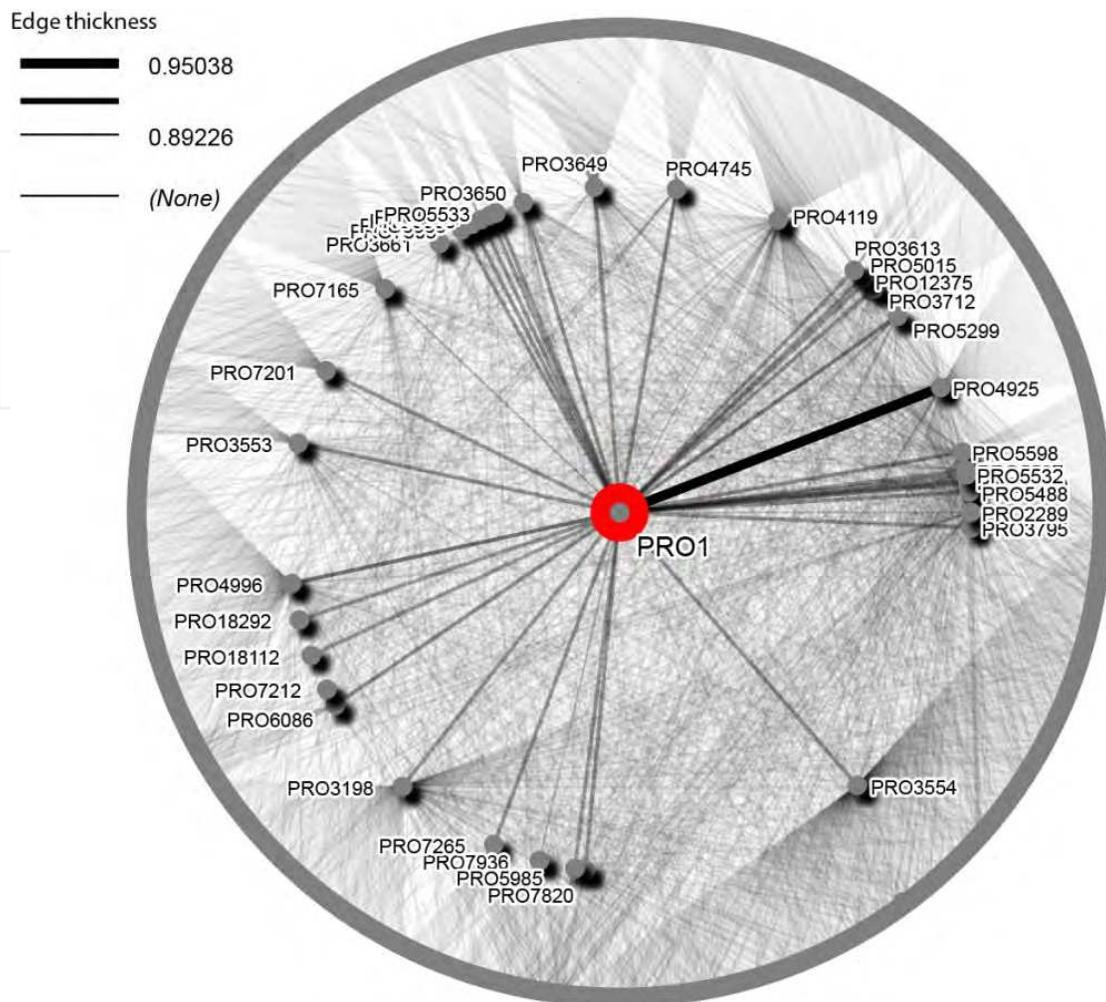


Fig. 4. Example graph laid out hierarchically, with PRO1 in a central position.

This is better, but still not that much more informative. One way of enriching our isolated data is by viewing it in the context of known and predicted interactions. I2D, the Interologous Interaction Database (<http://ophid.utoronto.ca/i2d>; (Brown et al., 2005, Brown et al., 2007)), will be our source for these interactions. NAViGaTOR offers an I2D plug-in, which enables the researcher to easily add interactions to the existing graph. NAViGaTOR also has the PSICQUIC search plug-in, which supports the searching of databases that implement the PSICQUIC interface (Aranda et al., 2011). To further support the openness and versatility of PPI integration, NAViGaTOR can import additional interactions from the same file formats listed above. If a database does not support any of these formats, finding or building a representation of the database in tab-delimited format may be an option as well. Our interaction search returns 1,367 nodes and 3,192 edges (Figure 3).

At this point, the graph has become more complex, and the force-directed layout is not helpful in interpreting it. Several options exist at this point for manually laying out objects in the graph. The user can select 'fix' nodes within the graph and either move them manually (which would be very labor intensive and inflexible) or lay them out with an array of tools such as linear, circular, arc or radial layout. We will use the radial layout method, starting with PRO1 as our central node and extending to a depth of 2. This gives us a hierarchical arrangement of

nodes starting with PRO1 in the centre, with its immediate interactors arranged circularly around it, and their interactors in turn arranged around them (Figure 4).

3.1 Ambiguity of protein names

When combining data from different sources, the users' choice of protein nomenclature becomes extremely important. Although a researcher knows which genes or proteins they are referring to, queries to a database require additional levels of specificity to resolve ambiguities in entity names.

For example, DLC1 has the following SwissProt identifiers: Q96QB1, Q9Y238, P63167, Q7Z5R8, Q45XF9, Q86UC6. However, names in literature could be ambiguous and confusing, potentially resulting in incorrect interpretation and analyses:

- **DLC1** (ARHGAP7) (KIAA1723) (STARD12) [**Rho GTPase-activating protein 7** (Rho-type GTPase-activating protein 7) (Deleted in liver cancer 1 protein) (Dlc-1) (StAR-related lipid transfer protein 12) (START domain-containing protein 12) (StARD12) (HP protein)]
- **DLEC1** (DLC1) [**Deleted in lung and esophageal cancer protein 1** (Deleted in lung cancer protein 1) (DLC-1)]
- **DYNLL1** (DLC1) (DNCL1) (DNCLC1) (HDLC1) [**Dynein light chain 1, cytoplasmic** (Dynein light chain LC8-type 1) (8 kDa dynein light chain) (DLC8) (Protein inhibitor of neuronal nitric oxide synthase) (PIN)]
- **DLC1** [**Deleted in liver cancer 1 variant 2** (Fragment)]
- **DLC1** [**DLC1 protein**]

Similarly, many papers refer to SHC - but details about which variant and which species are frequently "hidden" in the supplemental information (<http://www.cs.utoronto.ca/~juris/data/intech12/>). Yet, there are at least four variants in mouse and human. Sometimes, a radical change in nomenclature is required, such as in case of Caspases (Alnemri et al., 1986). Systematic analysis led to redefining various ICE, MACH, MCH genes into Caspase1-10 (Alnemri et al., 1986).

There are many different standards of referring to genes and proteins: UniProt (<http://www.uniprot.org>) (Jain et al., 2009), Ensembl (<http://www.ensembl.org>) (Flicek et al., 2012), EBI IPI (<http://www.ebi.ac.uk>) (Kersey et al., 2004), Gene Cards (<http://www.genecards.org>) (Safran et al., 2010), NCBI Gene (<http://www.ncbi.nlm.nih.gov>) (Maglott et al., 2010) are just a few examples of databases that attempt to systematically characterize and describe genes and proteins. Each database has its own focus and strengths, and different interaction or annotation databases may choose any one of these standards to organize their data. In this example, and in many other case uses of NAViGaTOR, the user may have to import data from one or more databases that use different nomenclatures. To facilitate the use of multiple nomenclatures, NAViGaTOR can store multiple IDs per node as a text feature, allowing alternative keys for node identification. When combining data from two or more databases using different formats, the user must translate between these different nomenclatures. This must be done very carefully and methodically, as this additional translation step often effects the data returned. For example, UniProt stores mappings from its own accession IDs to Ensembl

Gene IDs, and Ensembl stores mappings from its own IDs to UniProt. However, respectively, they return 55,639 unique UniProt accession IDs for 20,995 unique Ensembl gene IDs and 21,735 unique Ensembl gene IDs for 63,370 unique UniProt accession IDs. The mapping is clearly different depending on which method is used. There is no definitive mapping available in situations such as these: it is up to the individual researcher to choose and document the translations used to amalgamate their data in a fashion that is replicable. Bearing this in mind during the earlier stages of experiment design will make this process much easier and less prone to confusion or ambiguity.

3.2 Associating data with an existing graph

Though better organized, we still have in excess of 1,000 nodes and 3,000 interactions, and to better identify nodes and edges that represent novel research material, we must associate more data with those objects.

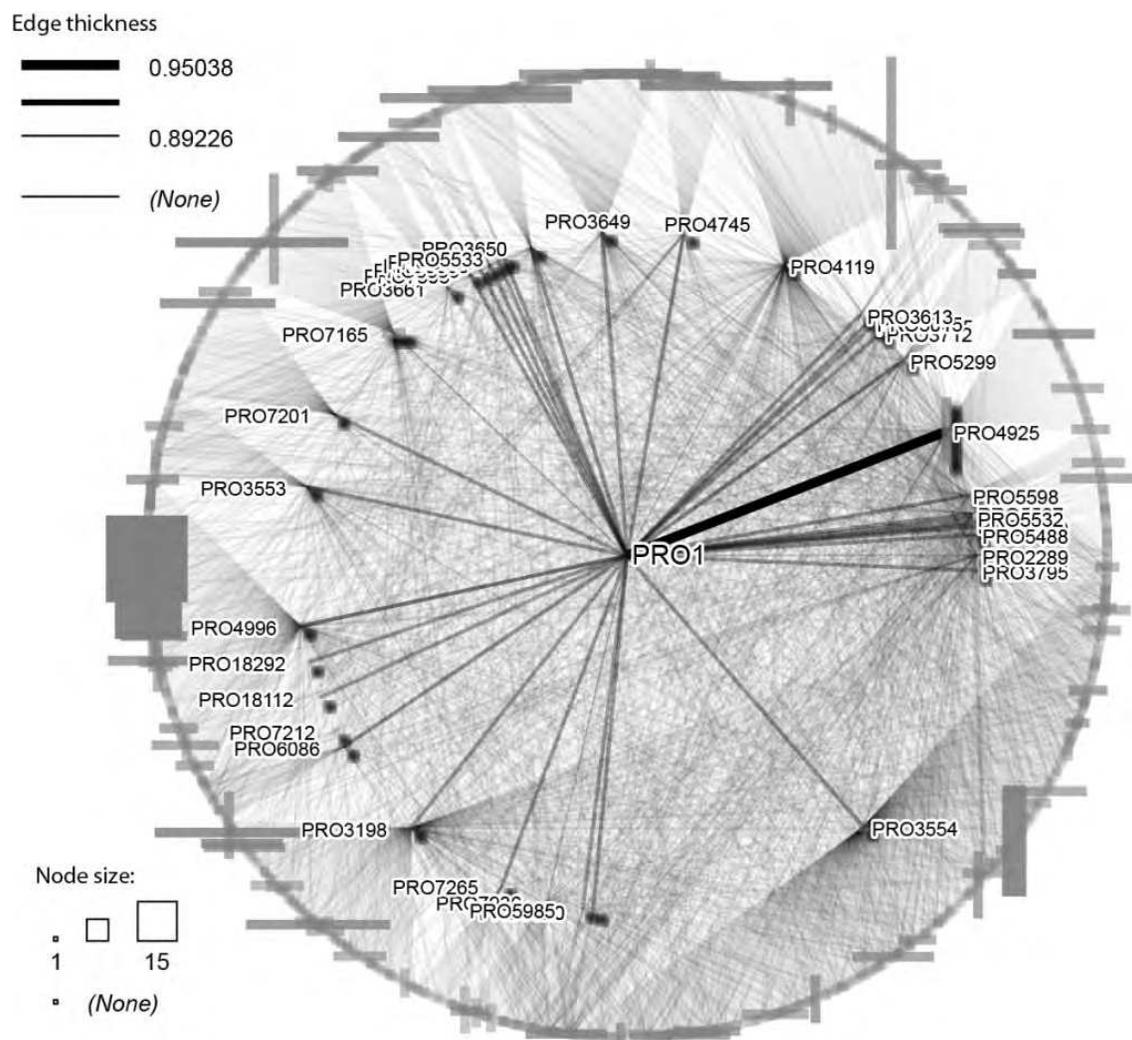


Fig. 5. Example graph with numbers of referencing studies in ovarian and prostate mapped to node width and height.

We can for example integrate PPIs with the gene expression results obtained from our literature studies. Each file contains several values associated with each gene, specifying the number of studies in which the gene was down-regulated, up-regulated and a total representing both (Figure 5). We will also generate a third file representing the total studies in which the gene was found to have been significantly deregulated, which simply sums the totals for the previous two files. Similarly to the opening of the initial experiment, NAViGaTOR requires a unique identifier column to be specified. In this case, because we are only concerned with data to be associated with nodes, the program only requires a single Node ID column. This process is the same for the prostate, ovarian cancer and generated data sets. To visualize this data, we will add another filter, this time mapping the total number of significantly deregulated studies in ovarian cancer to node width, and the total number of significantly deregulated studies in prostate cancer to its height. It is immediately evident which nodes have already been described to be up/down regulated in either one or both types of cancer. This can be useful to parallel the information already known from one cancer to the other. In addition, we can map the generated total of studies to node transparency, making genes with less disease evidence less obtrusive.

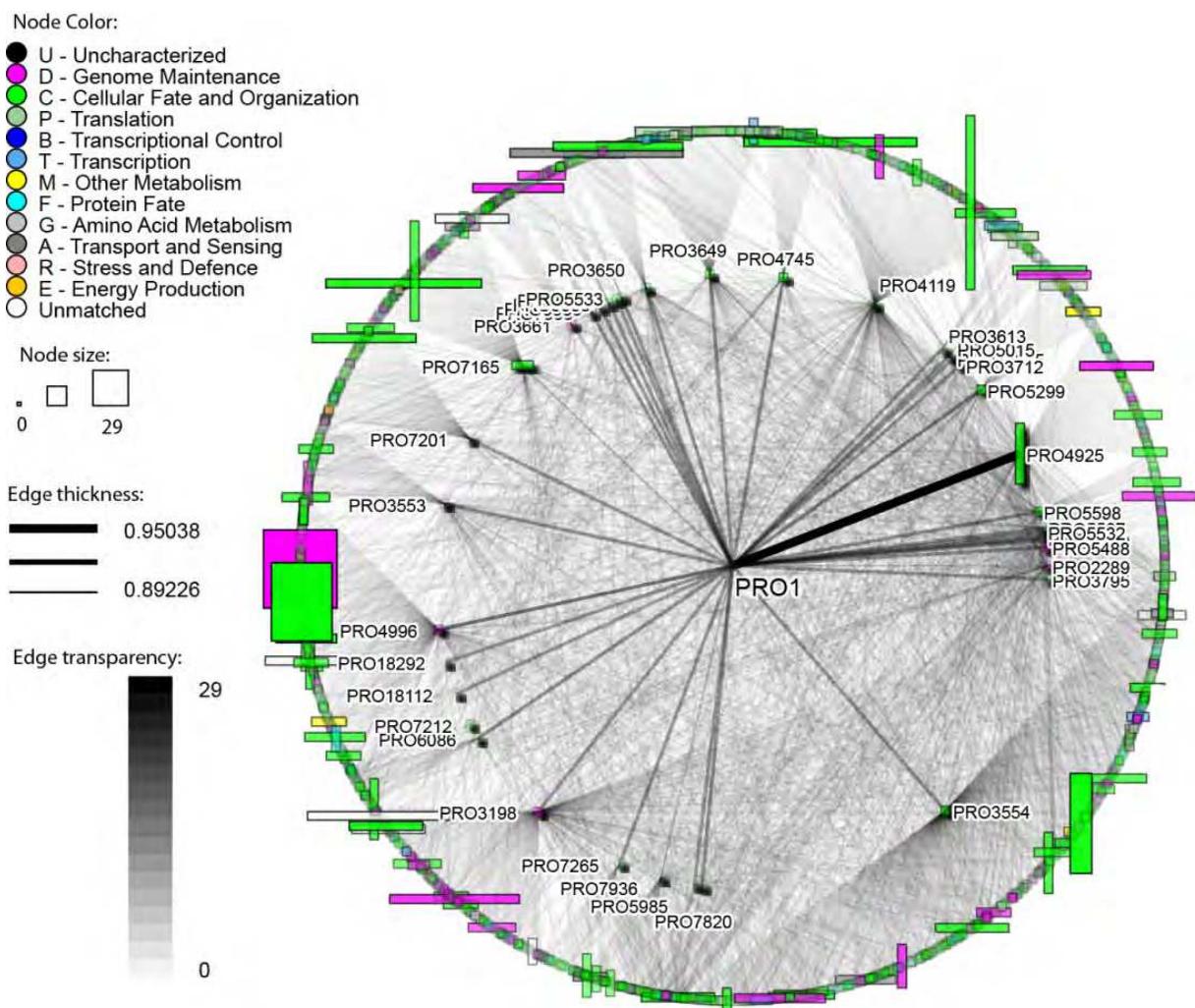


Fig. 6. Example graph with GO Annotation mapped to a color scheme.

We can also import structured data, in the form of GO attributes, retrieved from the I2D plug-in (Figure 6). We can view this data per individual node in the Node side panel, revealing the list of individual GO attributes and their descriptions. To get a graph-wide view of these attributes, we will add a filter to map the GO data to one of several categories, each with its own colour. The same result can be obtained by applying GO terms or other attributes, like pathways to which the node belongs, retrieved from other sources to the nodes as features and editing the filter in the desired way.

3.3 Importing drug-protein interactions

Finally, we will import a list of drugs and their gene targets as additional interactions. This expands our network to 2,707 nodes and 5,257 edges (Figure 7). Through a combination of manual layout and radial layout tools, we arrange the drugs in a circle around PRO1, its interactors, and their interactors from I2D. The edges connecting drugs to proteins are coloured blue to differentiate them from PPIs. To see the impact of individual drugs to this network, we map their degree to node size and transparency. Thus, large nodes represent drugs that target many of the proteins in the network. The top six of these drugs are labelled for convenience. Analogously, some proteins have a high degree of blue edges and connect to small nodes, such as ProX. These drugs show strong specificity to ProX. The initial data will be available in ASCII tab-delimited format and the final figure in NAViGaTOR 2 XML file at <http://www.cs.utoronto.ca/~juris/data/intech12/>.

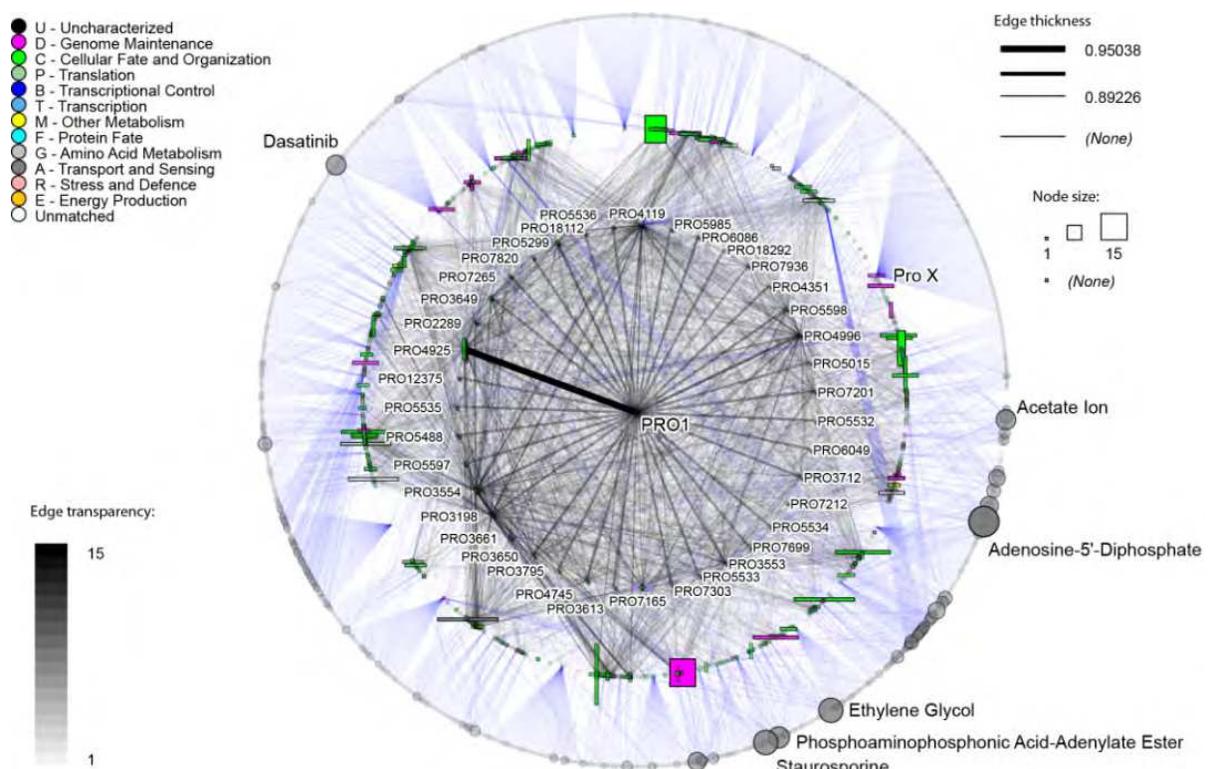


Fig. 7. Final graph, with drug interactions included and the size of nodes representing drugs derived from number of interactions within the graph. NAViGaTOR 2 XML file for the final figure is available in supplemental material (<http://www.cs.utoronto.ca/~juris/data/intech12/>).

4. Conclusions

Integrated databases and resources are only useful when they can be effectively accessed, navigated and analyzed. Several biological network visualization tools are currently available, providing a diverse range of approaches and algorithms. While many existing visualization tools are effective and widely used, there are several critical areas where these applications require improvement. Scalability is essential to visualize the tens of thousands of known PPIs, which is a challenge for current layout algorithms and software. Biological graph drawing software must also be able to handle richly annotated data, including genomic and proteomic profiles, pathways, Gene Ontology annotations and data in PSI-MI and BioPAX formats, in addition to the vast quantity of microarray and proteomic data that is available.

Individual tools need a good balance of performance and useful features. The features that are needed for each use are highly dependent on the available data and the workflow. As in any creative activity, a tool may enable new workflows by providing novel features, but the tool may also lack certain important features, or offer features that are not needed. There is no single solution that satisfies all of these requirements at the present time, and as data and workflows change over time, network visualization tools must also evolve.

As the data grow more complex, the performance of layout algorithms will need to improve, and new options of differentiating multiple attributes will be required. As certain workflows become more main-stream, they may be turned into *analysis patterns* and implemented as plug-ins. Standardizing file formats, APIs and plug-ins will further intertwine existing tools, enabling their easier integration and specialization.

With new data and advances in computational biology, user tasks are modified, which must be reflected by types of algorithms that support analyses and the user interfaces that effectively enable them. New graph theory algorithms for faster and biologically meaningful network layouts and algorithms for network structure analysis will need to be integrated into network visualization tools. Importantly, none of these algorithms would make a broad difference unless a user interface appropriate for biologists is available (Viau et al., 2010).

5. Acknowledgment

This research was funded in part by Ontario Research Fund (GL2-01-030), Canada Foundation for Innovation (CFI #12301 and CFI #203383), and the Ontario Ministry of Health and Long Term Care. The views expressed do not necessarily reflect those of the OMOHLTC. CP was funded in part by Friuli Exchange Program. IJ is supported in part by the Canada Research Chair Program.

The authors would like to thank Max Kotlyar, Dan Strumpf, Fiona Broackes-Carter and the entire Jurisica lab for useful comments and discussions.

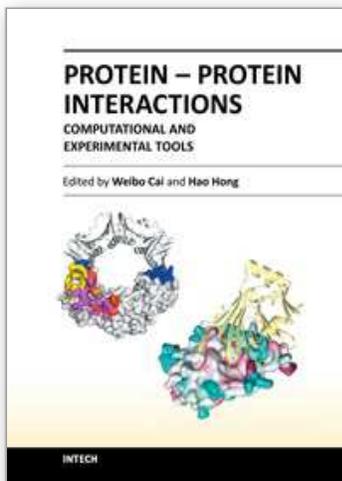
6. References

Alnemri, Emad, S., David J Livingston, Donald W Nicholson, Guy Salvesen, Nancy A Thornberry, Winnie W Wong, Junying Yuan (1986). Human ICE/CED-3 protease nomenclature, *Cell*, 87(2):171.

- Aranda, B., Blankenburg, H., Kerrien, S., Brinkman, F.S.L., Ceol, A., Chautard, E., Dana, J.M., De Las Rivas, J., Dumousseau, M., Galeota, E., Gaulton, A., Goll, J., Hancock, R.E.W., Isserlin, R., Jimenez, R.C., Kerssemakers, J., Khadake, J., Lynn, D.J., Michaut, M., O'Kelly, G., Ono, K., Orchard, S., Prieto, C., Razick, S., Rigina, O., Salwinski, L., Simonovic, M., Velankar, S., Winter, A., Wu, G., Bader, G.D., Cesareni, G., Donaldson, I.M., Eisenberg, D., Kleywegt, G.J., Overington, J., Ricard-Blum, S., Tyers, M., Albrecht, M., Hermjakob, H. (2011). PSICQUIC and PSIScore: Accessing and scoring molecular interactions. *Nature Methods*, 8(7): 28-529.
- Björn H. Junker, Christian Klukas and Falk Schreiber (2006). VANTED: A system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics*, 7:109
- Brown, K.R., and Jurisica, I. (2005). Online Predicted Human Interaction Database. *Bioinformatics*, 21(9):2076-82.
- Brown, K.R., and Jurisica, I. (2007). Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biol*, 8(5):R95.
- Brown, K.R., Otasek D, Ali M, McGuffin, M.J., Xie W, Devani B, van Toch I.L., and Jurisica I. (2009). NAViGaTOR: Network Analysis, Visualization and Graphing Toronto, *Bioinformatics*, 25(24): 3327-3329.
- Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, Wu G, D'Eustachio P, Schaefer C, Luciano J, Schacherer F, Martinez-Flores I, Hu Z, Jimenez-Jacinto V, Joshi-Tope G, Kandasamy K, Lopez-Fuentes AC, Mi H, Pichler E, Rodchenkov I, Splendiani A, Tkachev S, Zucker J, Gopinath G, Rajasimha H, Ramakrishnan R, Shah I, Syed M, Anwar N, Babur O, Blinov M, Brauner E, Corwin D, Donaldson S, Gibbons F, Goldberg R, Hornbeck P, Luna A, Murray-Rust P, Neumann E, Reubenacker O, Samwald M, van Iersel M, Wimalaratne S, Allen K, Braun B, Whirl-Carrillo M, Cheung KH, Dahlquist K, Finney A, Gillespie M, Glass E, Gong L, Haw R, Honig M, Hubaut O, Kane D, Krupa S, Kutmon M, Leonard J, Marks D, Merberg D, Petri V, Pico A, Ravenscroft D, Ren L, Shah N, Sunshine M, Tang R, Whaley R, Letovksy S, Buetow KH, Rzhetsky A, Schachter V, Sobral BS, Dogrusoz U, McWeeney S, Aladjem M, Birney E, Collado-Vides J, Goto S, Hucka M, Le Novère N, Maltsev N, Pandey A, Thomas P, Wingender E, Karp PD, Sander C, Bader GD. (2010). The BioPAX community standard for pathway data sharing. *Nature Biotechnology*. 28(9):935-42.
- Djebbari, A., Ali, M., Otasek, D., Kotlyar, M., Fortney, K., Wong, S., Hrvojić, A. and Jurisica, I. (2011). NAViGaTOR: Scalable and Interactive Navigation and Analysis of Large Graphs. *Internet Mathematics*, 7(4):314-347.
- Flicek P, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, Fitzgerald S, Gil L, Gordon L, Hendrix M, Hourlier T, Johnson N, Kähäri AK, Keefe D, Keenan S, Kinsella R, Komorowska M, Koscielny G, Kulesha E, Larsson P, Longden I, McLaren W, Muffato M, Overduin B, Pignatelli M, Pritchard B, Riat HS, Ritchie GR, Ruffier M, Schuster M, Sobral D, Tang YA, Taylor K, Trevanion S, Vandrovcova J, White S, Wilson M, Wilder SP, Aken BL, Birney E, Cunningham F, Dunham I, Durbin R, Fernández-Suarez XM, Harrow J, Herrero J, Hubbard TJ, Parker A, Proctor G, Spudich G, Vogel J, Yates A, Zadissa A, Searle SM. (2012). Ensembl. *Nucleic Acids Res.* [Epub ahead of print]

- Gehlenborg N., O'Donoghue S.I., Baliga N.S., Goesmann A., Hibbs M.A., Kitano H., Kohlbacher O., Neuweger H., Schneider R., Tenenbaum D., Gavin A.C. (2003). Visualization of omics data for systems biology. *Nat Methods*, 7(3 Suppl):S56-68.
- Helaers R, Bareke E, De Meulder B, Pierre M, Depiereux S, Habra N, Depiereux E. (2011). gViz, a novel tool for the visualization of co-expression networks. *BMC Res Notes*. 4(1):452.
- Himsolt, M. (1996). GML: A portable Graph File Format. Syntax. Retrieved from <http://www.fim.uni-passau.de/fileadmin/files/lehrstuhl/brandenburg/projekte/gml/gml-technical-report.pdf>
- Himsolt, M. (1996). GML: A portable Graph File Format. Syntax. Retrieved from <http://www.fim.uni-passau.de/fileadmin/files/lehrstuhl/brandenburg/projekte/gml/gml-technical-report.pdf>
- Hu Z., Hung J.H., Wang Y., Chang Y.C., Huang C.L., Huyck M., DeLisi C. (2009). VisANT 3.5: multi-scale network visualization, analysis and inference based on the gene ontology. *Nucleic Acids Res*, 37, W115–W121.
- Hu, Z., Mellor, J., Wu, J. and DeLisi, C. (2004). VisANT: an online visualization and analysis tool for biological interaction data. *BMC Bioinformatics*, 5, 17.
- Jain E., Bairoch A., Duvaud S., Phan I., Redaschi N., Suzek B.E., Martin M.J., McGarvey P., Gasteiger E. (2009). Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics*, 10:136.
- Junker B.H., Klukas C. & Schreiber F. (2006). VANTED: a system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics*, 7(109).
- Kerrien S, Orchard S, Montecchi-Palazzi L, Aranda B, Quinn AF, Vinod N, Bader GD, Xenarios I, Wojcik J, Sherman D, Tyers M, Salama JJ, Moore S, Ceol A, Chatr-Aryamontri A, Oesterheld M, Stümpflen V, Salwinski L, Nerothin J, Cerami E, Cusick ME, Vidal M, Gilson M, Armstrong J, Woollard P, Hogue C, Eisenberg D, Cesareni G, Apweiler R, Hermjakob H (2007). Broadening the horizon--level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol*. 5, 44
- Kersey P. J., Duarte J., Williams A., Karavidopoulou Y., Birney E., Apweiler R. (2004). The International Protein Index: An integrated database for proteomics experiments. *Proteomics* 4(7): 1985-1988
- Kreeger P.K., Lauffenburger D.A. (2010). Cancer systems biology: a network modeling perspective, *Carcinogenesis*, 31(1):2-8.
- Longabaugh WJ. (2012). BioTapestry: a tool to visualize the dynamic properties of gene regulatory networks. *Methods Mol Biol*. 786:359-94.
- Maglott D, Ostell J, Pruitt KD, Tatusova T. (2011). Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res*. 39(Database issue):D52-7.
- McGuffin, M, and Jurisica, I. (2009). Interaction techniques for selecting and manipulating subgraphs in network visualizations. *IEEE Trans Vis Comput Graph*, 15 (6): 937-944.
- Montoyo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, Morris Q, Bader GD (2010). GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics*, 26: 22
- Morrow JK, Tian L, Zhang S. (2010). Molecular networks in drug discovery. *Crit Rev Biomed Eng*. 38(2):143-56.

- Niu, Y., Otasek, D., Jurisica, I. (2011). Evaluation of linguistic features useful in extraction of interactions from PubMed; Application to annotating known, high-throughput and predicted interactions in I2D. *Bioinformatics*, 26(1): 111-9.
- Pavlopoulos G.A., Wegener A.L., Schneider R. (2008). A survey of visualization tools for biological network analysis, *BioData Min*, 1(12).
- Remmerie N., De Vijlder T., Laukens K., Dang T.H., Lemièrre F., Mertens I., Valkenborg D., Blust R., Witters E. (2011). Next generation functional proteomics in non-model plants: A survey on techniques and applications for the analysis of protein complexes and post-translational modifications. *Phytochemistry*, 72(10):1192-218.
- Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, Nativ N, Bahir I, Doniger T, Krug H, Sirota-Madi A, Olender T, Golan Y, Stelzer G, Harel A and Lancet D. (2010). GeneCards Version 3: the human gene integrator *Database*; doi: 10.1093/database/baq020
- Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D., Amin N., Schwikowski B., Ideker T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13:2498–2504.
- Shirdel EA, Xie W, Mak TW, Jurisica I. (2011) NAViGaTing the micronome--using multiple microRNA prediction databases to identify signalling pathway-associated microRNAs. *PLoS One*. 6(2):e17429.
- Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*. 1;27(3):431-2.
- Stein A., Mosca R., Aloy P. (2011). Three-dimensional modeling of protein interactions and complexes is going 'omics. *Curr Opin Struct Biol*, 21(2):200-8.
- Swainston N, Smallbone K, Mendes P, Kell D, Paton N. (2011). The SuBliMinaL Toolbox: automating steps in the reconstruction of metabolic networks. *J Integr Bioinform*. 8(2):186.
- The Gene Ontology Consortium. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet*. 25(1):25-9.
- Viau, C., McGuffin, M J., Chiricota, Y., and Jurisica, I. (2010). The FlowVizMenu and parallel scatterplot matrix: Hybrid multidimensional visualizations for network exploration. *IEEE Trans Vis Comput Graph*, 16(6):1100-8.
- Yu L.R. (2011). Pharmacoproteomics and toxicoproteomics: The field of dreams. *J Proteomics*, 74(12):2549-53.



Protein-Protein Interactions - Computational and Experimental Tools

Edited by Dr. Weibo Cai

ISBN 978-953-51-0397-4

Hard cover, 472 pages

Publisher InTech

Published online 30, March, 2012

Published in print edition March, 2012

Proteins are indispensable players in virtually all biological events. The functions of proteins are coordinated through intricate regulatory networks of transient protein-protein interactions (PPIs). To predict and/or study PPIs, a wide variety of techniques have been developed over the last several decades. Many in vitro and in vivo assays have been implemented to explore the mechanism of these ubiquitous interactions. However, despite significant advances in these experimental approaches, many limitations exist such as false-positives/false-negatives, difficulty in obtaining crystal structures of proteins, challenges in the detection of transient PPI, among others. To overcome these limitations, many computational approaches have been developed which are becoming increasingly widely used to facilitate the investigation of PPIs. This book has gathered an ensemble of experts in the field, in 22 chapters, which have been broadly categorized into Computational Approaches, Experimental Approaches, and Others.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

David Otasek, Chiara Pastrello and Igor Jurisica (2012). Scalable, Integrative Analysis and Visualization of Protein Interactions, Protein-Protein Interactions - Computational and Experimental Tools, Dr. Weibo Cai (Ed.), ISBN: 978-953-51-0397-4, InTech, Available from: <http://www.intechopen.com/books/protein-protein-interactions-computational-and-experimental-tools/scalable-integrative-analysis-and-visualization-of-protein-interactions>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen