BIOINFORMATICS

Vol. 21 Suppl. 2 2005, pages ii220–ii221
doi:10.1093/bioinformatics/bti1135

Systems Biology

Decomposing protein networks into domain–domain interactions

Mario Albrecht[1,*], Carola Huthmacher[1], Silvio C. E. Tosatto[2] and Thomas Lengauer[1]

[1]Max Planck Institute for Informatics, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany and [2]Department of Biology and CRIBI Biotechnology Center, University of Padova, Viale G. Colombo 3, 35121 Padova, Italy

ABSTRACT

Summary: The application of novel experimental techniques has generated large networks of protein–protein interactions. Frequently, important information on the structure and cellular function of protein–protein interactions can be gained from the domains of interacting proteins. We have designed a Cytoscape plugin that decomposes interacting proteins into their respective domains and computes a putative network of corresponding domain–domain interactions. To this end, the network graph of proteins has been extended by additional node and edge types for domain interactions, including different node and edge shapes and coloring schemes used for visualization. An additional plugin provides supplementary web links to Internet resources on domain function and structure.

Availability: Both Cytoscape plugins can be downloaded from http://www.cytoscape.org

Contact: mario.albrecht@mpi-inf.mpg.de

1 INTRODUCTION

Novel high-throughput techniques have generated large networks of protein–protein interactions, which need to be analyzed further using additional functional and structural data (Bork *et al*., 2004). Frequently, protein binding is characterized by specific interactions of evolutionarily conserved domains (Bornberg-Bauer *et al*., 2005). Important information on the cellular function of protein interactions and complexes can often be gained from the known functions of the interacting protein domains. Therefore, it is useful and often even necessary to decompose protein–protein interactions into their constituent domains before being able to functionally characterize them further and to model and investigate the spatial structure of protein complexes (Aloy *et al*., 2005).

In order to facilitate research on the molecular basis of an observed or predicted protein–protein interaction, we have designed a tool named DomainNetworkBuilder. It works as a Java plugin for Cytoscape, a free open-source software platform for the visualization and analysis of biomolecular networks (Shannon *et al*., 2003). This plugin DomainNetworkBuilder decomposes protein networks into domain–domain interactions and generates a new network of interacting domains. We have also implemented another Cytoscape plugin named DomainWebLinksPlugin that provides additional context-dependent web links to Internet resources on domain function and structure: databases of protein families, Pfam (Bateman *et al*., 2004), of interacting domains, InterDom (Ng *et al*., 2003), and of 3D interacting domains, 3did (Stein *et al*., 2005).

*To whom correspondence should be addressed.

2 MATERIALS AND METHODS

We have established a client–server architecture with the Cytoscape plugin DomainNetworkBuilder working as a client that queries an in-house MySQL database through our web server and processes the received data to create a network of interacting domains. The database stores synonyms for each gene/protein name, all protein domains from Pfam (Bateman *et al*., 2004), a special list of short repetitive Pfam domain motifs and domain–domain interactions with reliability scores from InterDom, a database of putatively interacting Pfam domains (Ng *et al*., 2003). It is possible to use other known or predicted domain–domain interactions alternatively or additionally to InterDom if a reliability score accompanies each interaction. Our database already covers all yeast proteins taken from UniProt (Bairoch *et al*., 2005), and it is currently being extended to human proteins and other species. A manually curated list of repetitive domain motifs was compiled based on the Pfam database field TP containing the keyword 'repeat'. This word indicates tandem sequence motifs such as HEAT or leucine-rich repeats forming one structural domain.

3 RESULTS AND DISCUSSION

After a protein network has been loaded as a graph consisting of nodes and edges, the DomainNetworkBuilder plugin can be executed in Cytoscape (Fig. 1, color version as online supplement). It uses the given protein labels in the network to retrieve the respective domain architectures and domain–domain interactions from our MySQL database. If a protein contains one or more domains, each domain is represented by a separate node labeled by the domain name and optionally by the protein name and by the start and end position of the domain in the respective protein sequence. The user can choose to disable the display of the respective protein nodes if domain nodes are available. If two or more proteins share the same name, one of the proteins is arbitrarily selected and a warning message is shown. Another message appears if the protein name is not found in the database. In this case, the protein will be handled as a protein without domains, and no domain nodes in addition to the protein node will be generated.

Like the interaction type 'pp' used by Cytoscape for a protein–protein interaction edge, we have introduced three new edge types for domain nodes: 'dl' for a domain linker between domain nodes of the same protein, 'pl' for a protein linker between a protein and domain node of the same protein, and 'dd' for a domain–domain interaction between different proteins. All domain nodes of the same protein are linearly connected by directed edges (arrows pointing

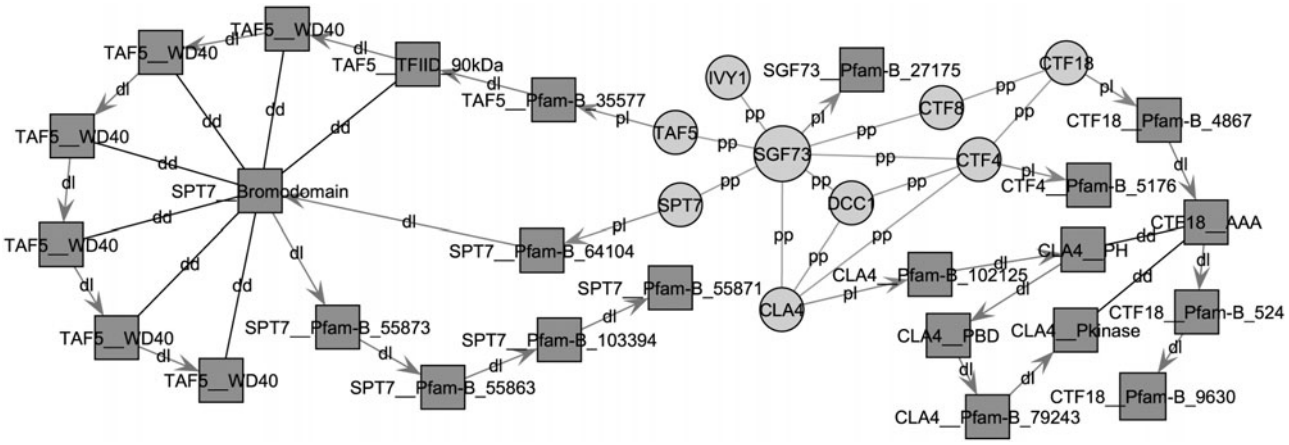

Fig. 1. Domain–domain interaction network around SGF73, the yeast homolog of ataxin-7 causative of the neurodegenerative disorder ataxia type 7 (Helmlinger *et al*., 2004). It is contained in transcriptional SAGA complexes that include TAF5 and SPT7 and show histone acetyltransferase activity. It may also play an important role in sister-chromatid cohesion, which involves an alternative replication factor C complex (CTF4, CTF8, CTF18 and DCC1) and presumably the protein kinase CLA4. Domain nodes are depicted as squares, protein nodes as circles. Edges are annotated by their respective interaction types.

from the N-terminus to the C-terminus). The user can choose whether this chain of domains is linked by a single directed edge to the protein node, which serves as N-terminal anchor, or each domain node belonging to a protein is connected directly to the protein node. The latter alternative may result in a closer local placement of the protein node to its domain nodes if appropriate graph drawing algorithms are applied.

Domain–domain interaction edges between different proteins are created only if the respective interaction score exceeds the overall threshold set by the user. If no domain–domain interaction edge can be established between two interacting proteins, the protein nodes remain connected. If more than one domain–domain interaction edge is possible between two proteins, the user can choose either always to select the edge between two domains with the maximum interaction score or to use all domain–domain edges (because two proteins could indeed interact through more than two domains).

Adjacent repetitive domain motifs constituting one structural domain need special treatment to avoid confusion of the network image. To select a subset from our manually curated list of ∼100 repetitive domain motifs of length up to ∼60, the user can set a threshold for the maximum motif length. All consecutive nodes of the same domain motif shorter than the threshold are merged into a single domain node. Further options offered to the user are that domain nodes without interactions to other proteins are not depicted and that Pfam-B domains can be ignored. Additional edge labels can consist of the interaction type or, in case of domain–domain interactions, of the interaction score. Moreover, the coloring schema as well as the different shapes of protein and domain nodes and interaction edges can easily be changed using the visualization tools of Cytoscape. The generated domain network can also be saved in file formats supported by Cytoscape.

4 CONCLUSIONS

Our Cytoscape plugin DomainNetworkBuilder provides tools for investigating and visualizing protein interactions on the more detailed molecular level of domains and binding sites. This approach assists in the validation and functional analysis of observed and predicted protein interactions, prioritization of further experiments, and 3D modeling of domain interactions and protein complexes.

ACKNOWLEDGEMENTS

M.A. has been supported by grants from the National Genome Research Network (NGFN) and the German Research Foundation (DFG) under contract number LE 491/14-1. S.T. has been funded by a ‘Rientro dei cervelli’ grant from the Italian Ministry of Education, University, and Research (MIUR). The research has been conducted in the context of the BioSapiens European Network of Excellence funded by the European Commission under grant number LSHG-CT-2003-503265.

Conflict of Interest: none declared.

REFERENCES

Aloy,P. *et al*. (2005) Protein complexes: structure prediction challenges for the 21(st) century. *Curr. Opin. Struct. Biol.*, **15**, 15–22.

Bairoch,A. *et al*. (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **33** (Database issue), D154–159.

Bateman,A. *et al*. (2004) The Pfam protein families database. *Nucleic Acids Res.*, **32**, D138–141.

Bork,P. *et al*. (2004) Protein interaction networks from yeast to human. *Curr. Opin. Struct. Biol.*, **14**, 292–299.

Bornberg-Bauer,E. *et al*. (2005) The evolution of domain arrangements in proteins and interaction networks. *Cell. Mol. Life Sci.*, **62**, 435–445.

Helmlinger,D. *et al*. (2004) Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes. *Hum. Mol. Genet.*, **13**, 1257–1265.

Ng,S.K. *et al*. (2003) InterDom: a database of putative interacting protein domains for validating predicted protein interactions and complexes. *Nucleic Acids Res.*, **31**, 251–254.

Shannon,P. *et al*. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.

Stein,A. *et al*. (2005) 3did: interacting protein domains of known three-dimensional structure. *Nucleic Acids Res.*, **33** (Database issue), D413–417.