



Article scientifique

Article

2011

Published version

Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

Snapshot: protein-protein interaction networks

Seebacher, Jan; Gavin Perrin, Anne-Claude

How to cite

SEEBACHER, Jan, GAVIN PERRIN, Anne-Claude. Snapshot: protein-protein interaction networks. In: Cell, 2011, vol. 144, n° 6, p. 1000–1000.e1. doi: 10.1016/j.cell.2011.02.025

This publication URL: <https://archive-ouverte.unige.ch/unige:154503>

Publication DOI: [10.1016/j.cell.2011.02.025](https://doi.org/10.1016/j.cell.2011.02.025)

SnapShot: Protein-Protein Interaction Networks

Jan Seebacher and Anne-Claude Gavin

Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL), 69117 Heidelberg, Germany

EXPERIMENTAL METHODS FOR CHARTING PROTEIN-PROTEIN INTERACTION NETWORKS

Binary interactions			Molecular machines/Protein complexes comembership	
Methods	Split proteins	Assay/Readout	Methods	
Yeast two-hybrid Protein fragment complementation assay	Transcription factor, ubiquitin Dehydrofolate reductase GFP or YFP	Transcription Antibiotic resistance Fluorescence	Affinity purification/Mass spectrometry Biochemical purification of affinity-tagged baits followed by MS identification of copurifying preys	

Protein pairs
+ Interacting
- Noninteracting

Interaction examples
- Signaling
- Enzyme - substrate

Interaction strength
- Transient to stable

Complex 1
Complex 2

Socio-affinity
- Allosteric
- Chaperone-assisted

Interaction strength
- Stable

INTERACTION NETWORK DATA QUALITY / BENCHMARK PARAMETERS

- "Gold standard" (PRS, positive reference list)
- Experimental PPI data set
- "Negative set" (RRS, random reference set, or set of proteins unlikely to interact)

PREDICTION OF PROTEIN COMPLEX TOPOLOGY

Spoke model

Matrix model

Socio-affinity model

Socio-affinity

NETWORK COMPONENTS

Party hubs: same time and space

Date hubs: different time and/or space

Hub: node with high degree

Edge: link between two nodes (interaction)

Node (protein)

Expression profiles and/or localization

NETWORK TOPOLOGIES

Random network

- Degrees follow Poisson (or peaked) distribution
- Vulnerability to failure

Scale-free network (Biological/cellular networks)

- Degrees follow power-law distributions
- Robustness against random failure
- Vulnerability to targeted attacks

Hierarchical network (Many types of real networks)

- Degrees follow power-law distributions
- Account for modularity, local clustering, and scale-free topology
- High clustering coefficient (C)

NETWORK MEASURES

Degree/ connectivity (k)	Clustering coefficient/ interconnectivity (C)	Assortativity/average nearest neighbor's connectivity (NC)	Shortest path (SP) between two nodes	Betweenness/ centrality (B)
<p>$k_A = \text{Nb of edges through } A = 5$</p>	<p>$C_A = \frac{\text{Actual links between A's neighbors (black)}}{\text{Possible links between A's neighbors (orange)}}$</p> <p>$C_A = n_A / [k_A(k_A - 1)/2]$ $= 2 / [4 \times (4 - 1) / 2] = 0.333$</p>	<p>$NC_A = (k_B + k_C + k_D + k_E + k_J) / 5$ $= (5 + 2 + 2 + 3 + 1) / 5 = 2.6$</p>	<p>$SP_{FH} = (F, D, A, B, H) = 4$</p>	<p>$B_A = \text{Fraction of SPs passing through } A = 0.090$</p>

SnapShot: Protein-Protein Interaction Networks

Cell

Jan Seebacher and Anne-Claude Gavin

Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL), 69117 Heidelberg, Germany

Cellular functions result from the coordinated action of groups of proteins interacting in molecular assemblies, pathways, or networks. The systematic and unbiased charting of protein-protein interaction (PPI) networks relevant to health or disease in a variety of organisms has become an important challenge in systems biology. Here, we review the main parameters and characteristics used to define PPI networks.

Experimental Methods for Charting Protein-Protein Interaction Networks

Genetic, *ex vivo* methods, such as the yeast two-hybrid system or protein complementation assays, measure pairwise binary, even transient, interactions. Biochemical approaches relying on affinity purification and mass spectrometry-based protein identification (AP-MS) are designed for the characterization of sturdy protein complexes. Direct, binary associations can be inferred from highly reciprocal AP-MS data sets by computational methods, such as the spoke and matrix models or socio-affinity scoring.

Data Quality

The overall quality of protein-protein interaction data sets is routinely evaluated through comparison with golden standards (lists of known high-confidence, literature-curated interacting proteins) and negative sets (lists of proteins that do not interact). This is generally replaced by sets of proteins that are unlikely to interact: for example, sets of random protein pairs or lists of proteins never reported to “co-occur” (coexpression, colocalization, coconservation, gene fusion, synteny, text mining, etc.). To ensure proper use by other scientists, researchers are encouraged to follow the molecular interaction experiment (MIMIX) guidelines when reporting PPI data to public databases (Orchard et al., 2007).

The following benchmark parameters can be deduced: (1) the false positive rate (FPR, background) is the fraction of identified interactions that are in the negative set; (2) the true positive rate (TPR, or accuracy) is the fraction of identified, genuine interacting protein pairs: $TPR = 1 - FPR$; (3) the false negative rate (FNR) is the fraction of the golden standard interactions that have been missed; (4) the coverage is the fraction of the golden standard covered by an experimental data set: $coverage = 1 - FNR$; and (5) the true negative rate (TNR) is the fraction of unidentified, genuine, noninteracting protein pairs.

Prediction of Protein Complex Topologies

The spoke model assumes that the bait interacts directly with each one of the copurifying proteins. The matrix model assumes that any two proteins within the set of copurifying proteins have pairwise direct interactions. The socio-affinity scoring integrates both the spoke and matrix models. It measures the frequency with which pairs of proteins were found associated within sets of biochemical purifications; protein pairs with high scores are more likely to be in direct physical contact.

Network Components

Node, protein; edge, interaction or link between two nodes; hub, node with a high degree (see below); date hub, hub that does not coexpress and/or colocalize with interacting nodes (interactions at different times or locations); party hub, hub that coexpresses and/or colocalizes with interacting nodes (simultaneous interactions).

Network Topologies and Structures

In random networks, all nodes have approximately the same number of edges (same degree, see below). In scale-free networks, most of the nodes have only a few edges, and a few nodes (also called hubs) have a very large number of edges. The degree distribution approximates a power law. In hierarchical networks, sparsely linked nodes are components of highly clustered areas, and a few hubs connect these highly clustered regions. Hierarchical networks are characterized by their modularity, a power-law degree distribution, and a large average clustering coefficient. The discrete topologies confer different network vulnerability to perturbation. Whereas random networks are sensitive to random attacks, scale-free networks are resistant to such failures. However, they show susceptibility to targeted attacks.

Network Measures

The degree, or connectivity (k), measures how many edges a node has to other nodes. The degree distribution, $P(k)$, gives the probability that a selected node has exactly k edges. The degree distribution is used to characterize different classes of networks. The clustering coefficient (C) measures the degree of interconnectivity in the neighborhood of a node. This is the ratio between the number of observed links between the neighbors of a node and the number of all possible links between the neighbors of this node if all of this node's neighbors were connected to each other. The average clustering coefficient, $\langle C \rangle$, characterizes the overall tendency of nodes to form clusters. It is a measure of a network's hierarchical property. The assortativity, or neighbor connectivity (NC), represents the average degree of the nearest neighbors of a node. In PPI networks, highly connected nodes (hubs) do not link directly to each other but interact with sparsely connected nodes. The distance, or path length, between two nodes represents the number of edges that separate the two nodes. There are many alternative paths between two nodes; the shortest path (SP) has the smallest number of edges. The betweenness (B) is a measure of a node's centrality in a network. It represents the fraction of all of the shortest paths between all nodes in a network that pass through a given node.

Protein-Protein Interaction Resources

Biocarta, www.biocarta.com; BioGRID, www.thebiogrid.org; BOND, www.bond.unleashedinformatics.com; DIP, www.dip.doe-mbi.ucla.edu; HPRD, www.hprd.org; IntAct, www.ebi.ac.uk/intact; MINT, www.mint.bio.uniroma2.it; Reactome, www.reactome.org; SGD, www.yeastgenome.org; STRING, www.string-db.org.

Visualization Tools for Biological Networks

Cytoscape, www.cytoscape.org; NAVIGATOR, www.ophid.utoronto.ca/navigator; Osprey, www.biodata.mshri.on.ca/osprey/servlet/Index.

REFERENCES

- Behrends, C., Sowa, M.E., Gygi, S.P., and Harper, J.W. (2010). Network organization of the human autophagy system. *Nature* 466, 68–76.
- Han, J.D.J., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P., and Vidal, M. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 430, 88–93.
- Kühner, S., van Noort, V., Betts, M.J., Leo-Macias, A., Batisse, C., Rode, M., Yamada, T., Maier, T., Bader, S., Beltran-Alvarez, P., et al. (2009). Proteome organization in a genome-reduced bacterium. *Science* 326, 1235–1240.
- Orchard, S., Salwinski, L., Kerrien, S., Montecchi-Palazzi, L., Oesterheld, M., Stümpflen, V., Ceol, A., Chatr-aryamontri, A., Armstrong, J., Woollard, P., et al. (2007). The minimum information required for reporting a molecular interaction experiment (MIMIX). *Nat. Biotechnol.* 25, 894–898.
- Tarassov, K., Messier, V., Landry, C.R., Radinovic, S., Serna Molina, M.M., Shames, I., Malitskaya, Y., Vogel, J., Bussey, H., and Michnick, S.W. (2008). An *in vivo* map of the yeast protein interactome. *Science* 320, 1465–1470.
- Venkatesan, K., Rual, J.F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.I., et al. (2009). An empirical framework for binary interactome mapping. *Nat. Methods* 6, 83–90.
- Yamada, T., and Bork, P. (2009). Evolution of biomolecular networks: lessons from metabolic and protein interactions. *Nat. Rev. Mol. Cell Biol.* 10, 791–803.
- Yu, H., Braun, P., Yildirim, M.A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., et al. (2008). High-quality binary protein interaction map of the yeast interactome network. *Science* 322, 104–110.