

Reviews in Computational Biology:

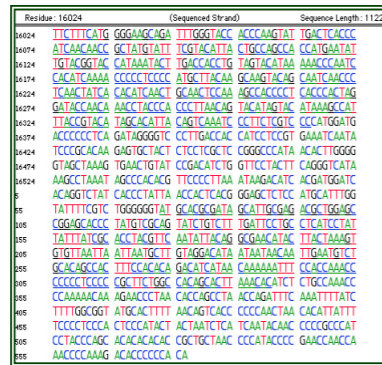
**Can biophysical properties of PPIs
improve the network descriptions of
metabolism?**

James Smith

Department of Biochemistry,
University of Cambridge

History of Network Biology

Sequence Analysis



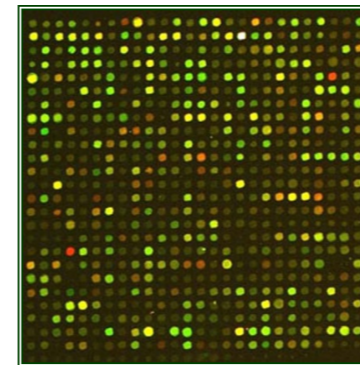
- Gene sequencing
- Sequence alignments
- Homologue searches
- Sequence motif finding

Structure Analysis



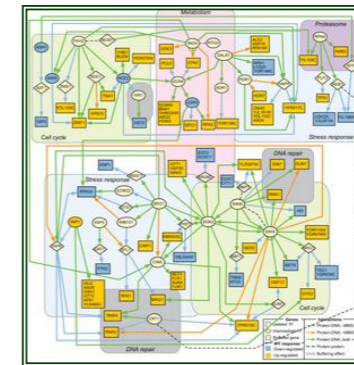
- Homologue searches
- Protein folding
- Binding site prediction
- Functional predictions

Expression Analysis



- Function prediction
- Gene clustering
- Sample classification

Network Analysis



- Function prediction
- Pathway identification
- Module detection
- Network modelling

Bioinformatics

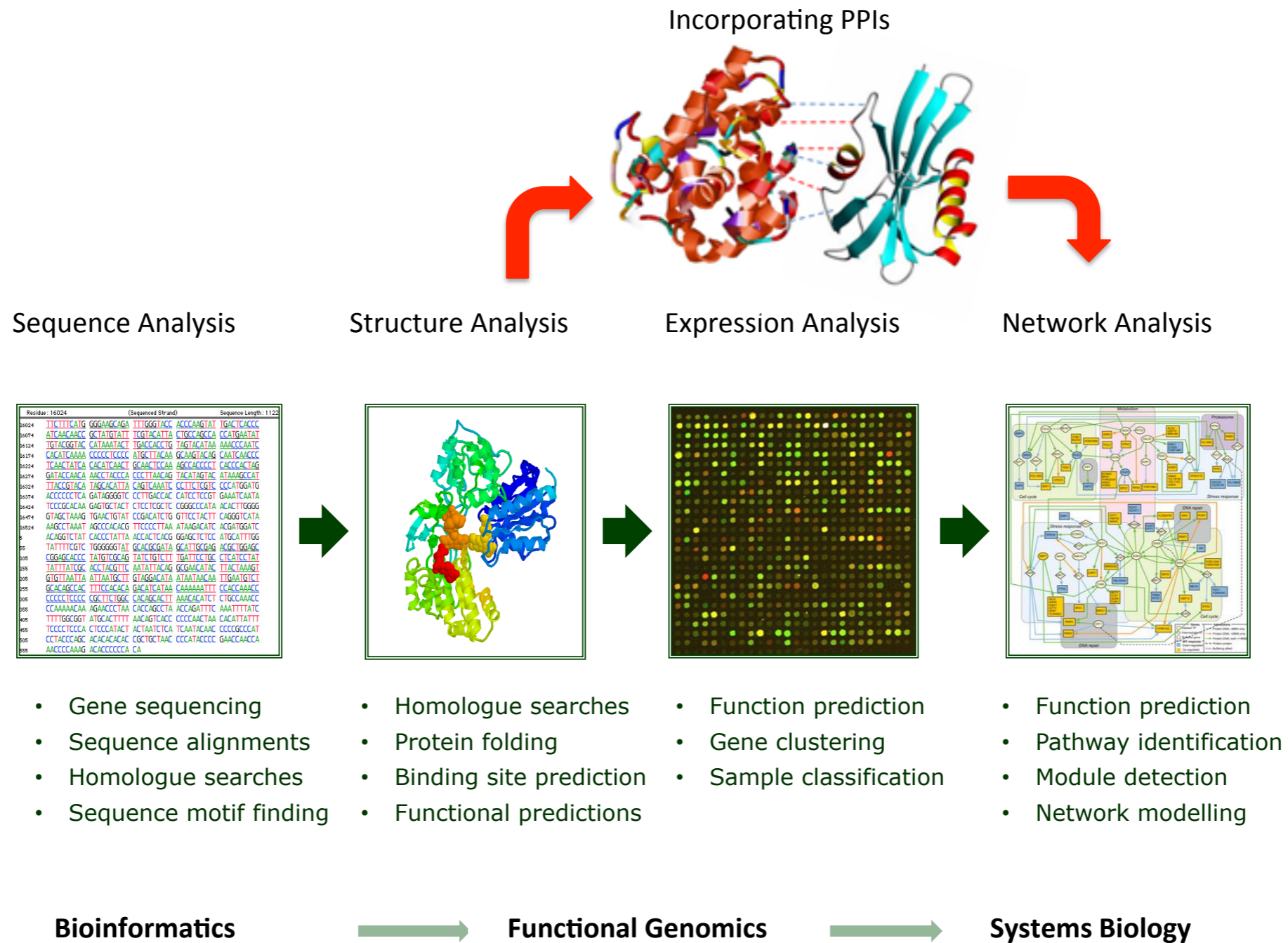


Functional Genomics

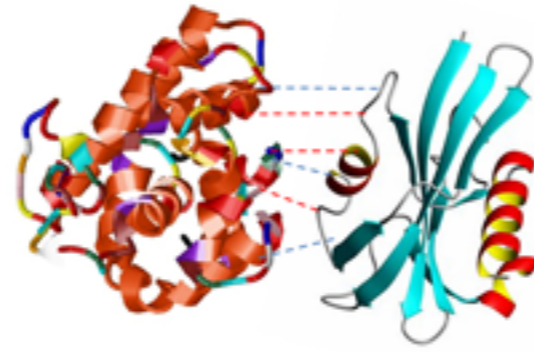


Systems Biology

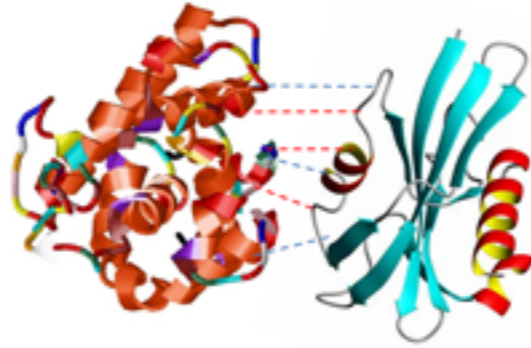
History of Network Biology



1. Protein-Protein Interactions

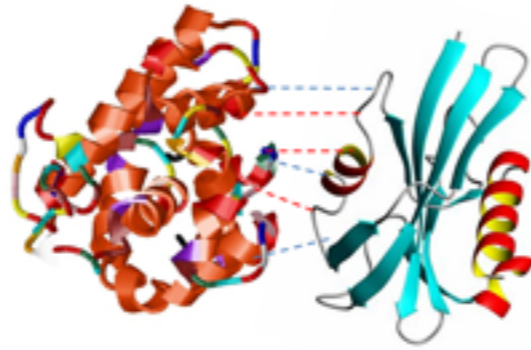


1. Protein-Protein Interactions

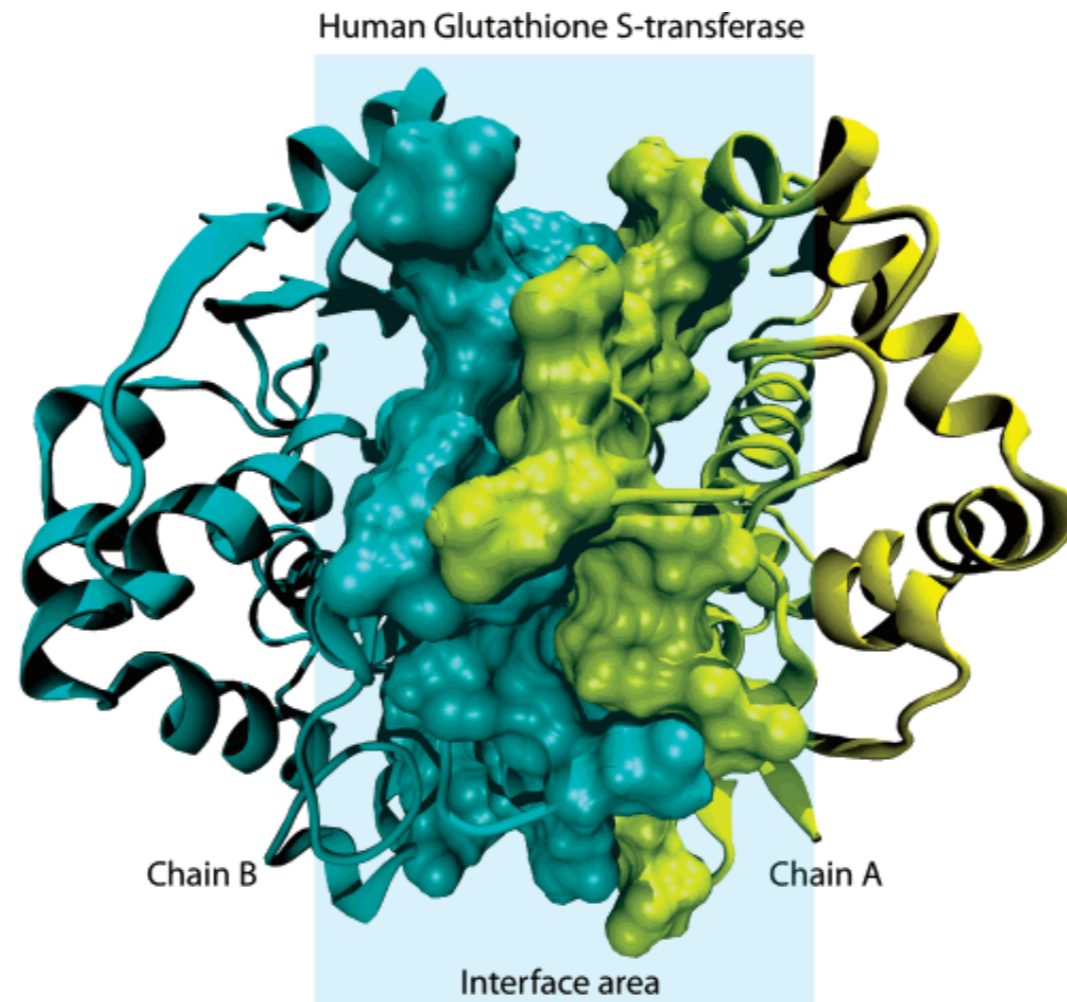


Is there a precise definition within the range of biomolecular interactions?

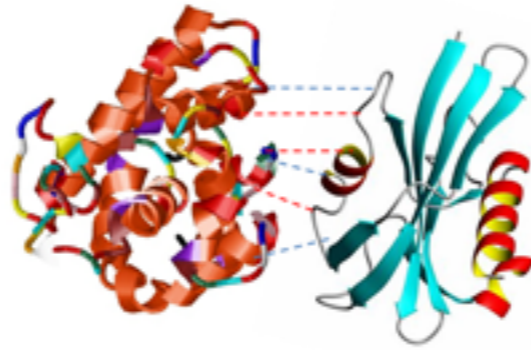
1. Protein-Protein Interactions



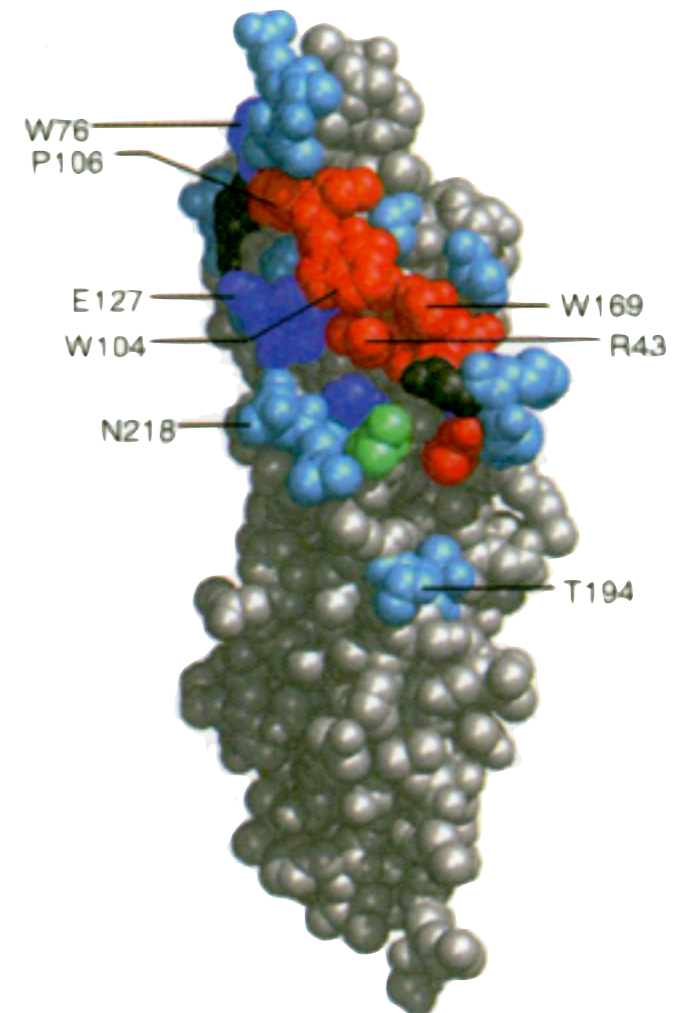
- Protein-pairs, direct physical contact, specific-distinct interface, functional role



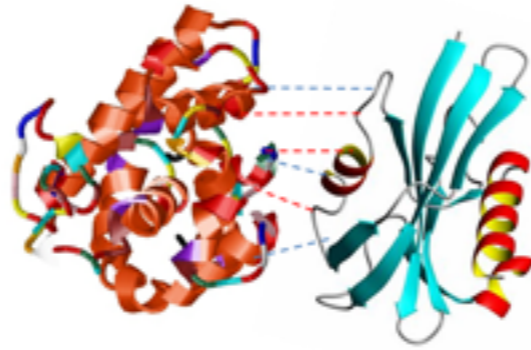
1. Protein-Protein Interactions



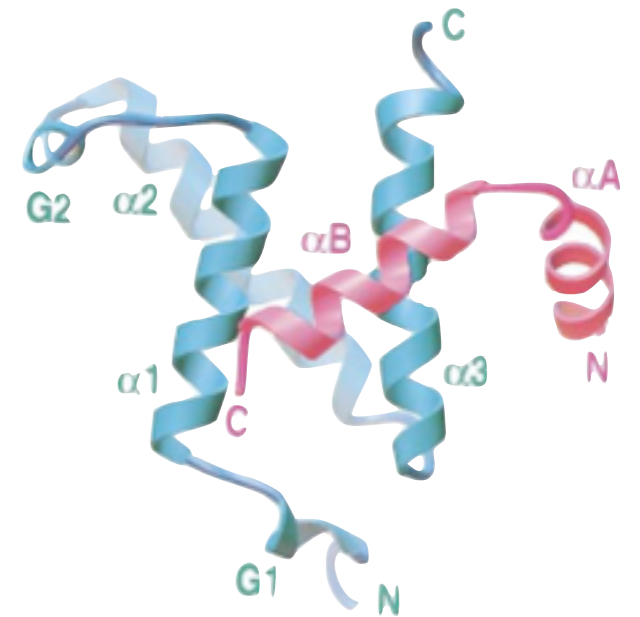
- Protein-pairs, direct physical contact, specific-distinct interface, functional role
- Surface complementarity, well-defined interface, includes clusters of residues



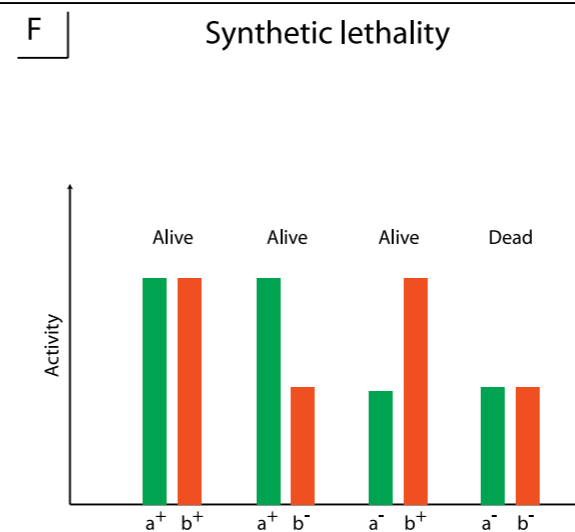
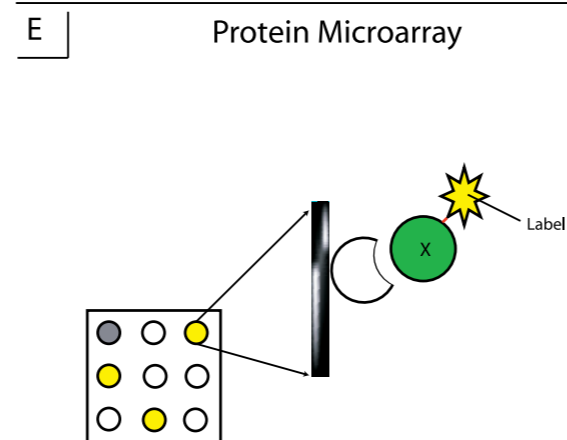
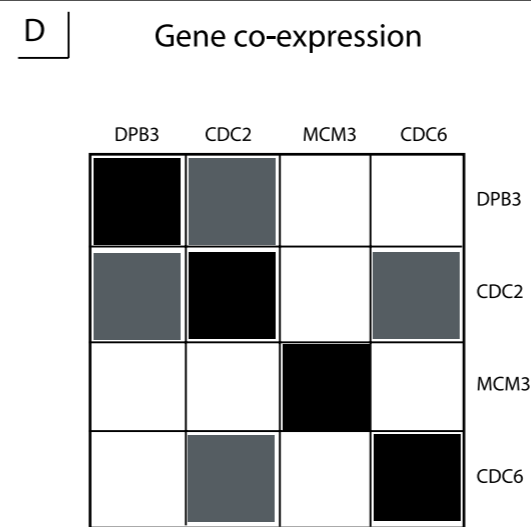
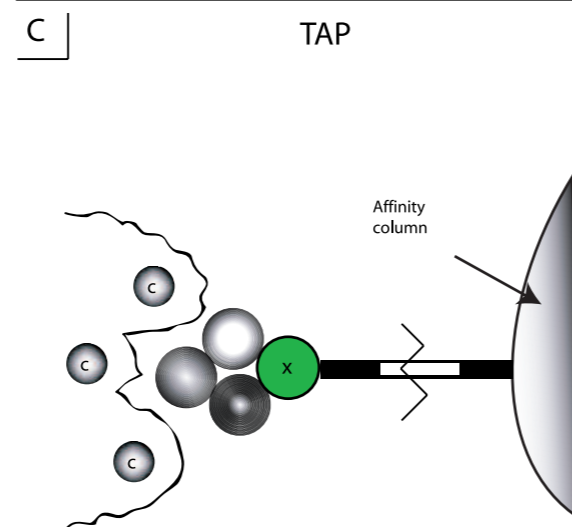
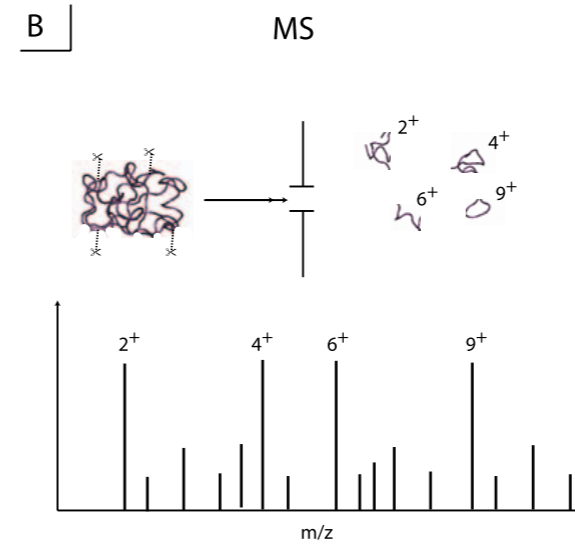
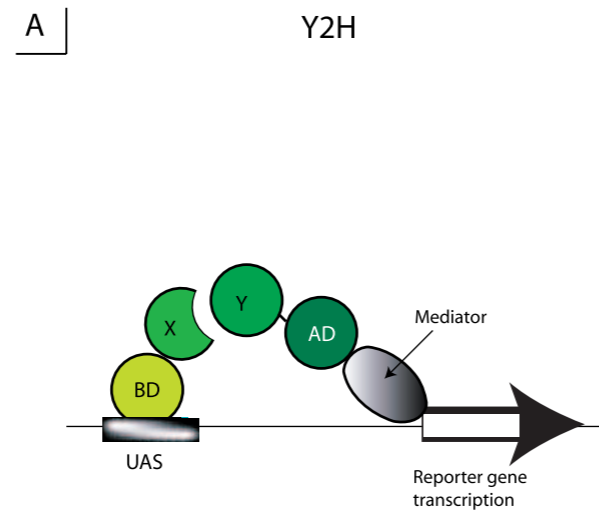
1. Protein-Protein Interactions



- Protein-pairs, direct physical contact, specific-distinct interface, functional role
- Surface complementarity, well-defined interface, includes clusters of residues
- Concerted binding of SSEs, localised disordered to ordered transitions of SSEs

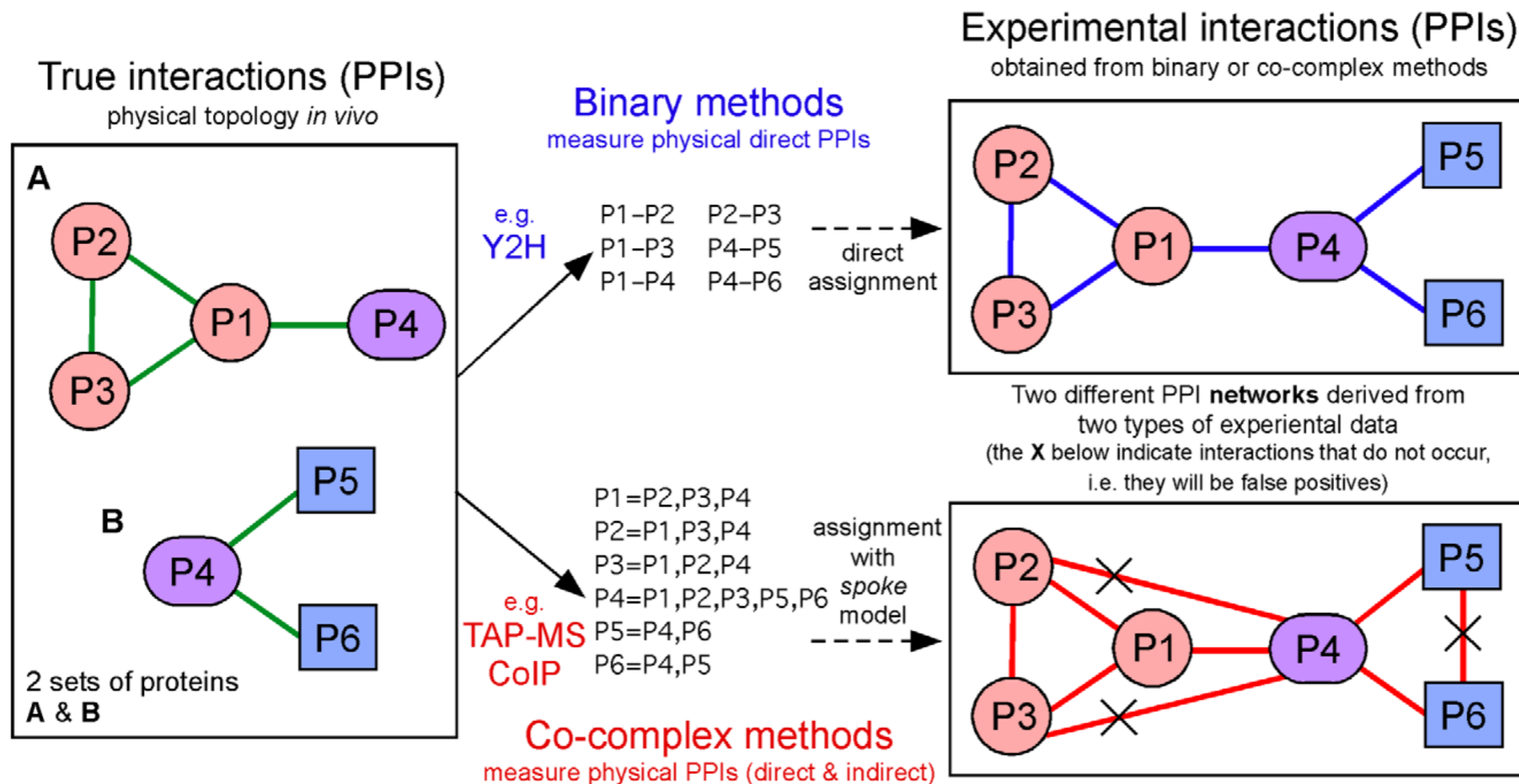


Methods to identify PPIs



Methods to identify PPIs

Most widely used is the combination of Yeast Two-Hybrid with Tandem Affinity Purification coupled with Mass Spectrometry



Methods to Predict PPIs

- Rosetta Stone
infers protein linkage from
genomic analyses
- Phylogenetic Profile
identifies genes that are
correlated across genomes
- Conserved Gene Neighbour
identifies proximal genes across genomes
- Operon/Gene Cluster
can assign putative function to unknown genes

Method Name	Protein/Domain Interaction	Physical Interaction/ Functional Association
Gene co-expression	P	F
Synthetic lethality	P	F
Gene cluster and gene neighbor	P	F
Phylogenetic profile	P, D	F
Rosetta Stone	P	F
Sequence co-evolution	P, D	F
Classification	P, D	P
Integrative	P, D	P
Domain association	D	P
Bayesian networks	P, D	F, P
Domain pair exclusion	D	P
p -Value	D	P

Second column shows if method is designed to predict protein (P) or domain (D) interactions (note that predicted domains can also be used for verifying protein interactions).

Third column shows if the method can be used to infer direct physical interaction (P) or indirect functional association (F).

doi:10.1371/journal.pcbi.0030043.t001

Only half can infer physical associations

Example Databases of PPIs

The image displays three overlapping browser windows showcasing biological databases for protein-protein interactions (PPIs).

- Top Left Window (Pathguide):** Shows the Pathguide website with a navigation menu on the left and a main content area titled "Complete Listing of All Pathguide Resources".
- Top Right Window (MINT):** Shows the MINT (Molecular Interaction Network Tool) database homepage. It features a navigation bar with "Home", "Search", "Curation", "Statistics", "Download", and "Contacts/Links/Linking". A statistics box indicates 239353 interactions, 34685 proteins, and 5139 pmids. A welcome message states: "Welcome to MINT, the Molecular Interaction database. MINT focuses on experimentally verified protein-protein interactions mined from the scientific literature by expert curators. The full MINT dataset can be freely downloaded." There are also news items about UniProt API and Palsic query results.
- Bottom Window (ConsensusPathDB):** Shows the ConsensusPathDB homepage from the Max Planck Institute for Molecular Genetics. It features the CPDB logo and navigation links for "home", "content information", "search", "interactions of molecules/pathways", "shortest interaction paths", "gene set functional analysis", "over-representation analysis", "enrichment analysis", "network upload and mapping", "download / data access", "documentation", and "news".

ConsensusPathDB-human integrates functional interaction networks in *Homo sapiens* including binary and complex protein protein, genetic, metabolic, signaling, gene regulatory and drug-target interactions, as well as biochemical pathways. Data originate from currently 29 public resources for functional interactions (listed below) and interactions that we have curated from the literature. The interaction data are integrated in a complementary manner (avoiding redundancies), resulting in a seamless interaction network containing different types of interactions.

If you use ConsensusPathDB, please cite:

Kamburov, A. *et al.* (2011) ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Res.* 39:D712-717.
 Pentchev, K. *et al.* (2010) Evidence mining and novelty assessment of protein-protein interactions with the ConsensusPathDB plugin for Cytoscape. *Bioinformatics* 26:2796-2797.
 Kamburov, A. *et al.* (2009) ConsensusPathDB—a database for integrating human functional interaction networks. *Nucleic Acids Res.* 37:D623-628.

Integrated databases:

name	version	protein	signaling	metabolic	gene	genetic	drug-target	biochemical
unique physical entities:		51,425						
unique functional interactions:		169,918						
pathways:		3,281						

At the bottom of the ConsensusPathDB window, there are sections for "NEWS" (Announcements about the most recent additions and changes to the database), "REGISTRATION/ACCOUNT" (Registration and account maintenance), "STATISTICS" (Detailed information about the current state of the database), "SATELLITES" (DIP-related projects), and "SERVICES" (DIP-derived services).

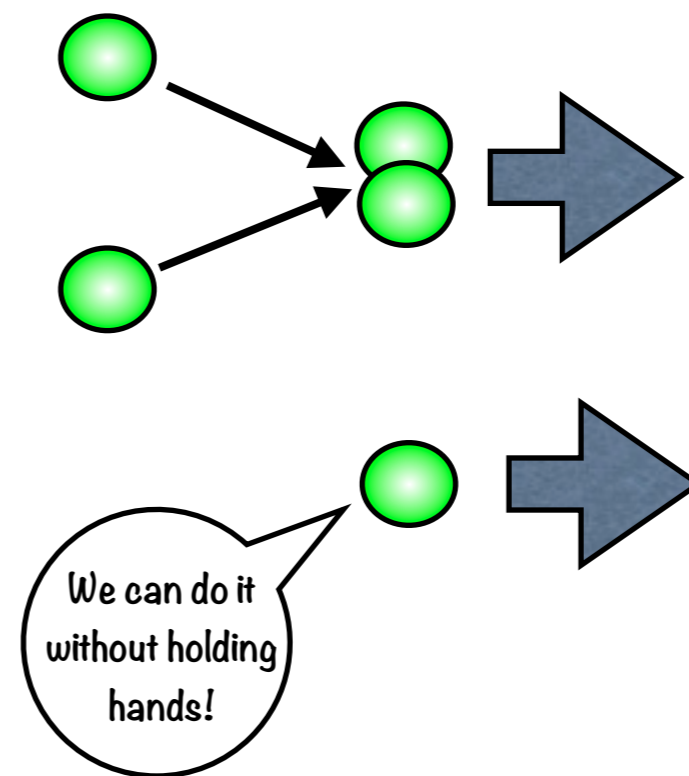
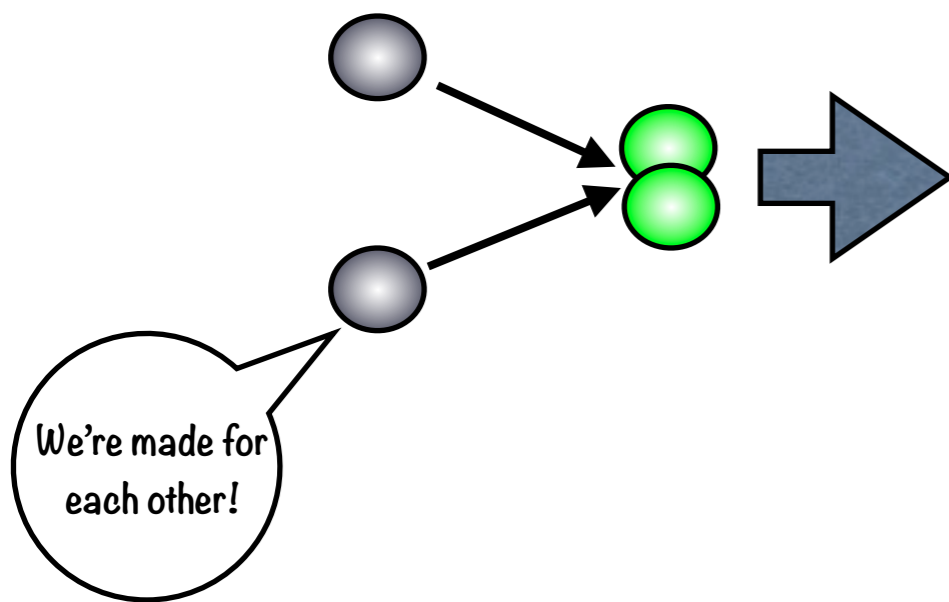
2. Biophysics of PPIs

Biophysics of the PPI complex

- **Most detail comes from structural biology**

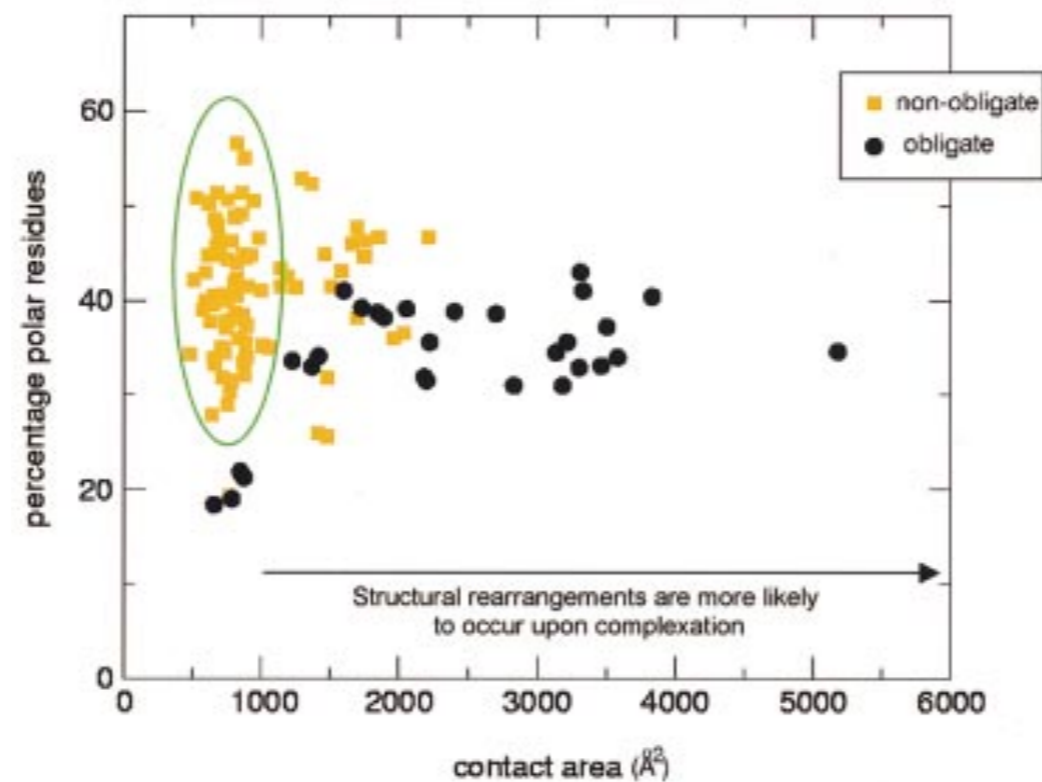
Biophysics of the PPI complex

- Most detail comes from structural biology
- Obligatory and Non-obligatory Complexes



Biophysics of the PPI complex

- Most detail comes from structural biology
- Obligatory and Non-obligatory Complexes



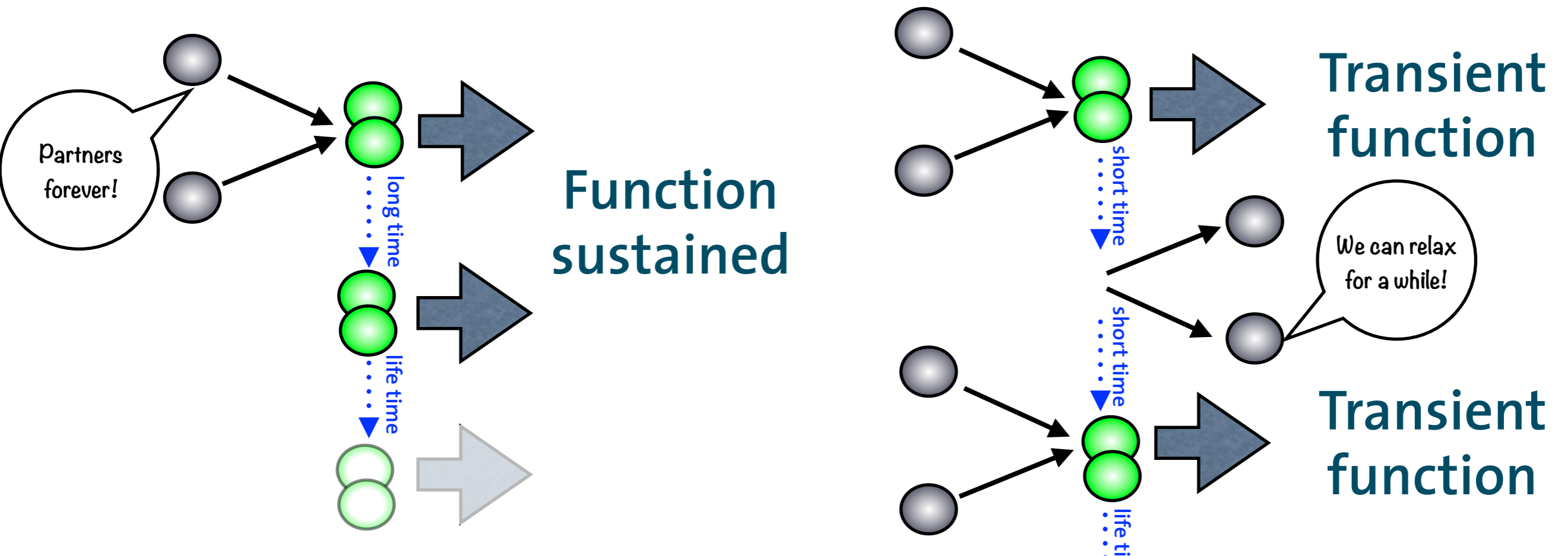
Contact area and polarity of the interfaces of various non-obligate and obligate complexes. Obligatory complexes with a small and hydrophobic interface include coiled-coil proteins. The ellipse denotes the contact area–polarity space of weak transient interactions.

Biophysics of the PPI complex

- **Most detail comes from structural biology**
- **Obligatory and Non-obligatory Complexes**
- **Permanent or Transient Complexes**

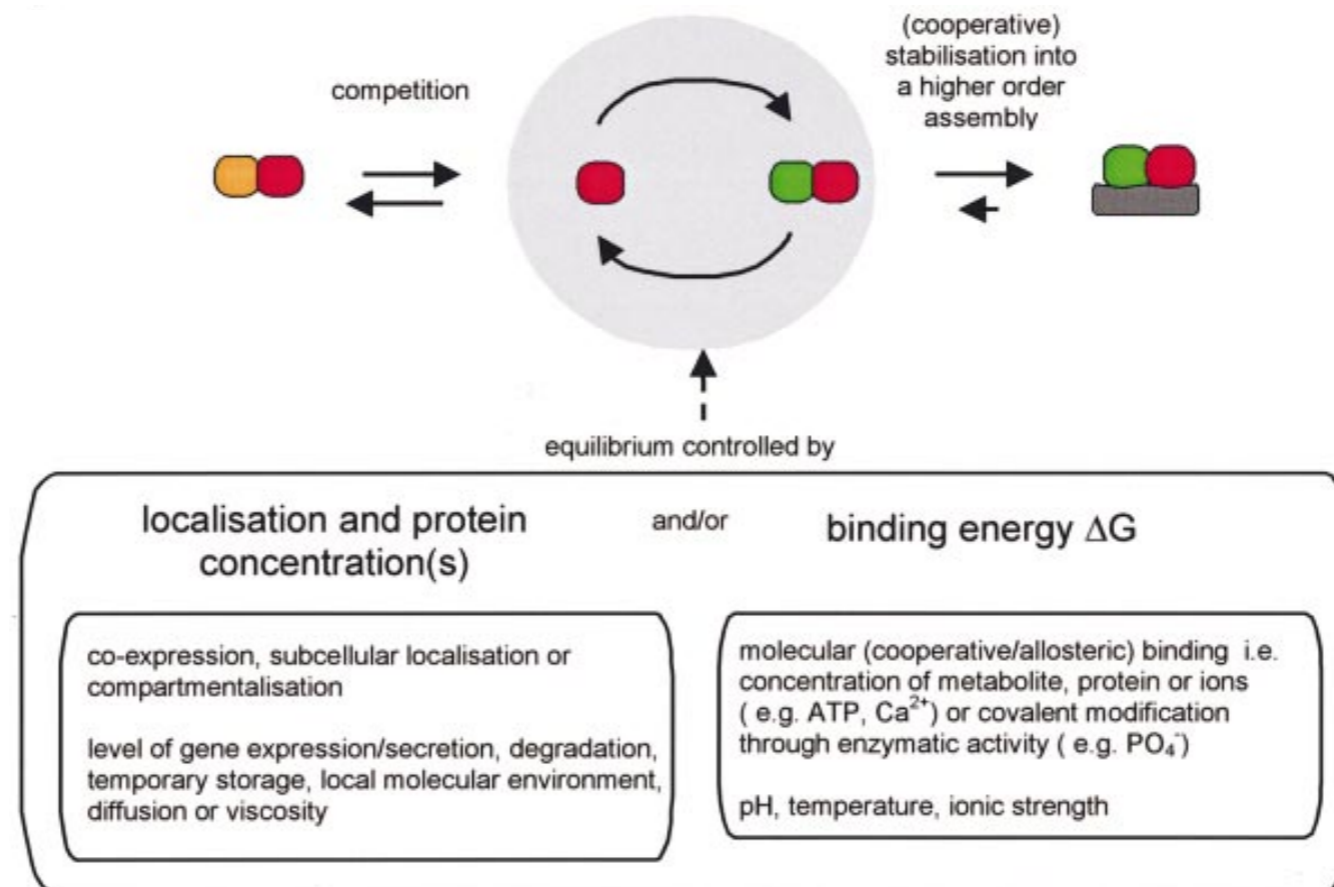
Biophysics of the PPI complex

- Most detail comes from structural biology
- Obligatory and Non-obligatory Complexes
- Permanent or Transient Complexes

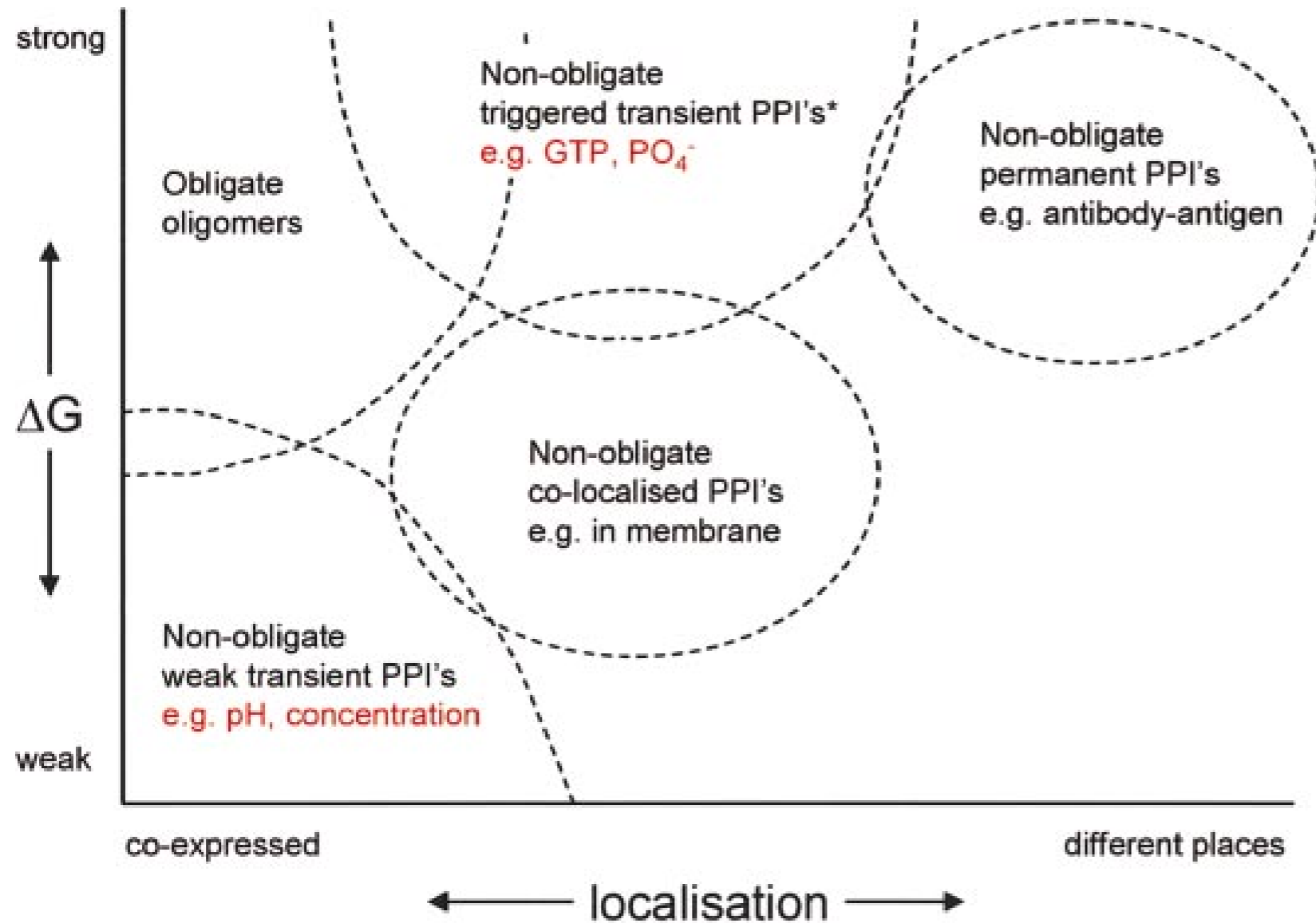


Biophysics of the PPI complex

- Most detail comes from structural biology
- Obligatory and Non-obligatory Complexes
- Permanent or Transient Complexes



Biophysics of the PPI complex



The protein networks found underline the multi-specificity and dynamics of PPIs involving transient interactions, since many proteins were shown to be involved in more than one multiprotein complex or binary interaction. These multicomponent transient complexes have yet to be characterized in terms of their detailed structures or energetics. We can expect the full range of interactions, from rigid to dynamic, weak to strong, obligate and non-obligate. The functional rationale for many of these complexes is not known, and ultimately the optimization of function during evolution will be the key determinant of the observed character of each complex.

Thoughts on the Biophysical Properties of Complexation of Transient PPIs?

What's happening at the molecular level (temporally and spatially)?

Thoughts on the Biophysical Properties of Complexation of Transient PPIs?




- Small scale: on-rates
affinity driven - long range attractions
eg electrostatics, solvation and ionic milieu
- Small scale: off-rates
min destabilising but max stabilising
intermolecular contact area
- Small scale: reordering and stabilising
secondary & tertiary structure

Studies on the Biophysical Properties of Complexation of Transient PPIs?

Current Opinion in Structural Biology

Volume 10, Issue 2, 1 April 2000, Pages 153–159

Electrostatic aspects of protein–protein interactions

Felix B Sheinerman , Raquel Norel , Barry Honig 



Transient protein–protein interactions

Protein Engineering, Design and Selection (2011) 24(9): 635–648 first published online June 15, 2011

PNAS

Free-energy distribution of binary protein–protein binding suggests cross-species interactome differences

Yi Y. Shi*, Gerald A. Miller[†], Hong Qian*, and Karol Bomsztyk^{‡§}

Departments of *Applied Mathematics, [†]Physics, and [‡]Medicine, University of Washington, Seattle, WA 98195

Communicated by Ernest M. Henley, University of Washington, Seattle, WA, May 25, 2006 (received for review April 19, 2006)

Published online 4 June 2010

Nucleic Acids Research, 2010, Vol. 38, Web Server issue W407–W411
doi:10.1093/nar/gkq502

ANCHOR: a web server and database for analysis of protein–protein interaction binding pockets for drug discovery

Lidio M. C. Meireles¹, Alexander S. Dömling^{2,3,*} and Carlos J. Camacho¹

Biological Protein-Protein Interaction Prediction Using Binding Free Energies and Linear Dimensionality Reduction

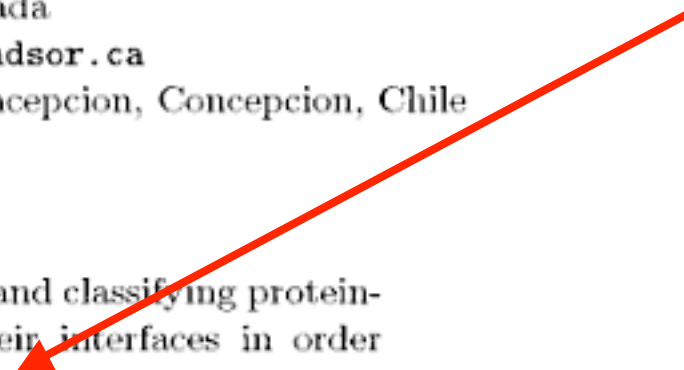
L. Rueda¹, Carolina Garate², Sridip Banerjee¹, and Md. Mominul Aziz¹

¹ School of Computer Science, University of Windsor, 401 Sunset Ave.,
Windsor, ON, N9B 3P4, Canada
{lrueda,banerje1,azizc}@cs.uwindsor.ca

² Department of Computer Science, University of Concepcion, Concepcion, Chile
cgarate@udec.cl

Abstract. An important issue in understanding and classifying protein-protein interactions (PPI) is to characterize their interfaces in order to discriminate between transient and obligate complexes. We propose a classification approach to discriminate between these two types of complexes. Our approach uses contact and binding free energies of the residues present in the interaction, which are the input features for the classifiers. A total of 282 features are extracted for each complex, and the classification is performed via recently proposed dimensionality reduction (LDR) methods, including the well-know Fisher's discriminant analysis and two heteroscedastic approaches. The results on a standard benchmark of transient and obligate protein complexes show that LDR approaches achieve a very high classification accuracy (over 78%), outperforming various support vector machines and nearest-neighbor classifiers. An additional insight on the proposed approach and experiments on different subsets of features shows that solvation energies can be used in the classification, leading to a performance comparable to using the full binding free energies of the interaction.

Defined as a continuum
not considering localisation



Thoughts on the Biophysical Properties of Complexation of PPIs?

What's happening at a larger scale of
organisation?

Thoughts on the Biophysical Properties of Complexation of PPIs?

- Large scale: self-organisation around cellular architecture (cytoskeleton, organelles)
- Large scale: transcriptional changes during the cell cycle or under nutrient stress

Example References

Physical Biology > Volume 2 > Number 2

Ozlem Keskin *et al* 2005 *Phys. Biol.* **2** S24 doi:10.1088/1478-3975/2/2/S03

Protein–protein interactions organization, cooperativity and mapping in a bottom-up Systems Biology approach

Ozlem Keskin^{1,2}, Buyong Ma², Kristina Rogale³, K Gunasekaran² and Ruth Nussinov^{2,4}

The large-scale organization of metabolic networks

H. Jeong*, B. Tombor†, R. Albert*, Z. N. Oltvai† & A.-L. Barabási*

NATURE | VOL 407 | 5 OCTOBER 2000 | www.nature.com

Nature **444**, 383-386 (16 November 2006) | doi:10.1038/nature05201; R 4 September 2006; Published online 15 October 2006

Visualization of transient encounter complexes in protein–protein association

Chun Tang¹, Junji Iwahara¹ & G. Marius Clore¹

1. Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, USA

Correspondence to: G. Marius Clore¹ Correspondence and requests for materials should be addressed to G.M.C. (Email: mariusc@intra.niddk.nih.gov).

Protein–protein interaction networks: how can a hub protein bind so many different partners?
Trends in Biochemical Sciences, Volume 34, Issue 12, December 2009, Pages 594-600
Chung-Jung Tsai, Buyong Ma, Ruth Nussinov [View Abstract](#)

BMC Systems Biology


BioMed Central

Research article

Open Access

The integrated analysis of metabolic and protein interaction networks reveals novel molecular organizing principles

Pawel Durek* and Dirk Walther

3. Graph and Network representations

Graph and Network representations

Algorithms for Molecular Biology



Research

Open Access

Decomposition of overlapping protein complexes: A graph theoretical method for analyzing static and dynamic protein associations

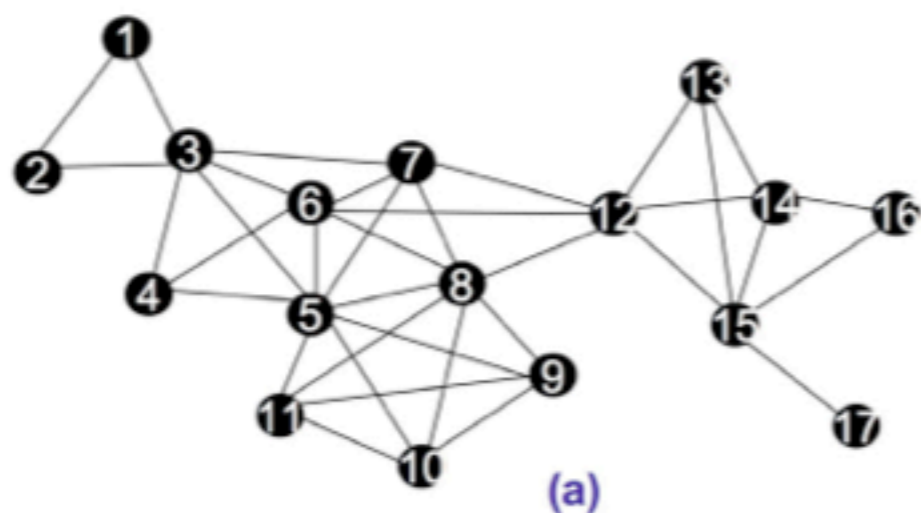
Elena Zotenko^{1,2}, Katia S Guimarães^{1,3}, Raja Jothi¹ and Teresa M Przytycka^{*1}

Abstract

Background: Most cellular processes are carried out by multi-protein complexes, groups of proteins that bind together to perform a specific task. Some proteins form stable complexes, while other proteins form transient associations and are part of several complexes at different stages of a cellular process. A better understanding of this higher-order organization of proteins into overlapping complexes is an important step towards unveiling functional and evolutionary mechanisms behind biological networks.

Results: We propose a new method for identifying and representing overlapping protein complexes (or larger units called *functional groups*) within a protein interaction network. We develop a graph-theoretical framework that enables automatic construction of such representation. We illustrate the effectiveness of our method by applying it to TNF α /NF- κ B and pheromone signaling pathways.

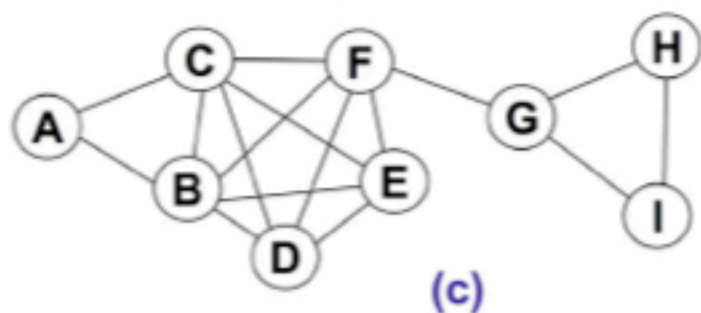
Conclusion: The proposed representation helps in understanding the transitions between functional groups and allows for tracking a protein's path through a cascade of functional groups. Therefore, depending on the nature of the network, our representation is capable of elucidating temporal relations between functional groups. Our results show that the proposed method opens a new avenue for the analysis of protein interaction networks.



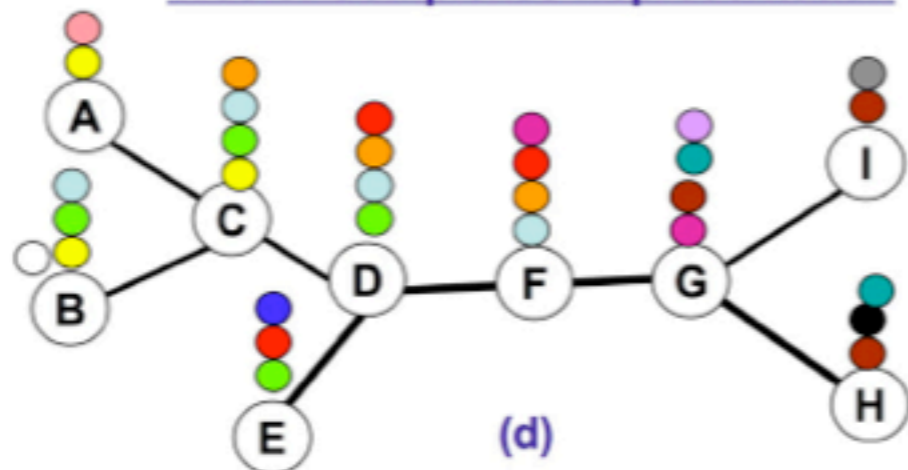
Maximal Cliques

- A 1,2,3
- B 3,4,5,6
- C 3,5,6,7
- D 5,6,7,8
- E 5,8,9,10,11
- F 6,7,8,12
- G 12,13,14,15
- H 14,15,16
- I 15,17

Naïve Maximal Clique Overlap Representation



Tree of Complexes Representation



- | | |
|-------|-----------|
| ● 1,2 | ● 9,10,11 |
| ● 3 | ● 12 |
| ○ 4 | ● 13 |
| ● 5 | ● 14 |
| ○ 6 | ● 15 |
| ● 7 | ● 16 |
| ● 8 | ● 17 |

(e)

Figure 2

A Hypothetical Protein Interaction Network. (a) A hypothetical protein interaction network. (b) A list of all maximal cliques in the network. (c) A naive representation of overlaps between maximal cliques. Each maximal clique is a node and there is an edge between two maximal cliques if and only if they share a protein. (d) The clique tree representation. Once again, every maximal clique is a node, but the cliques are connected in such a way that the resulting graph is a tree. Moreover, cliques that contain a given protein form a connected subgraph. (e) This color scheme is used to show the subtree of every

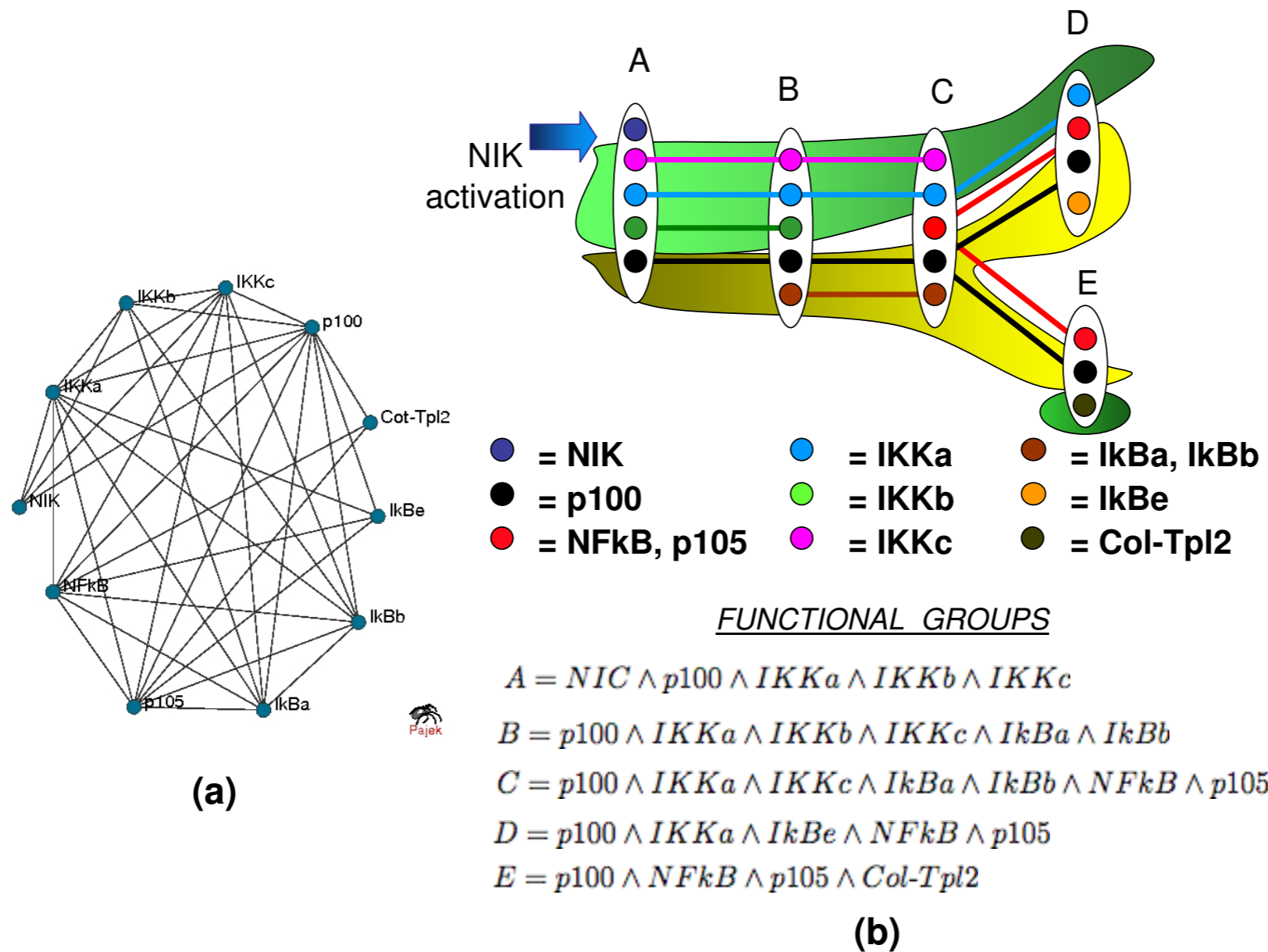


Figure 3
TNF α /NF- κ B Signaling Pathway. The TNF α /NF- κ B signaling pathway. (a) The network. (b) The Tree of Complexes representation. The flow of action is visually represented by background colors: green for activators (IKKs) and yellow for inhibitors (I κ Bs, and p100). The NIK kinase is in the first functional group (A), together with all three members of the IKK complex and p100. Functional group B includes, in addition to p100, the IKKs and two inhibitors I κ B α and I κ B β . This group is the beginning of interaction between IKKs and I κ Bs. Functional group C loses some of the IKKs, continues to show I κ B and begins to show interaction between I κ Bs and NF- κ B factors. Finally, in group E we see the entrance of NIK-independent Col-Tpl2 kinase.

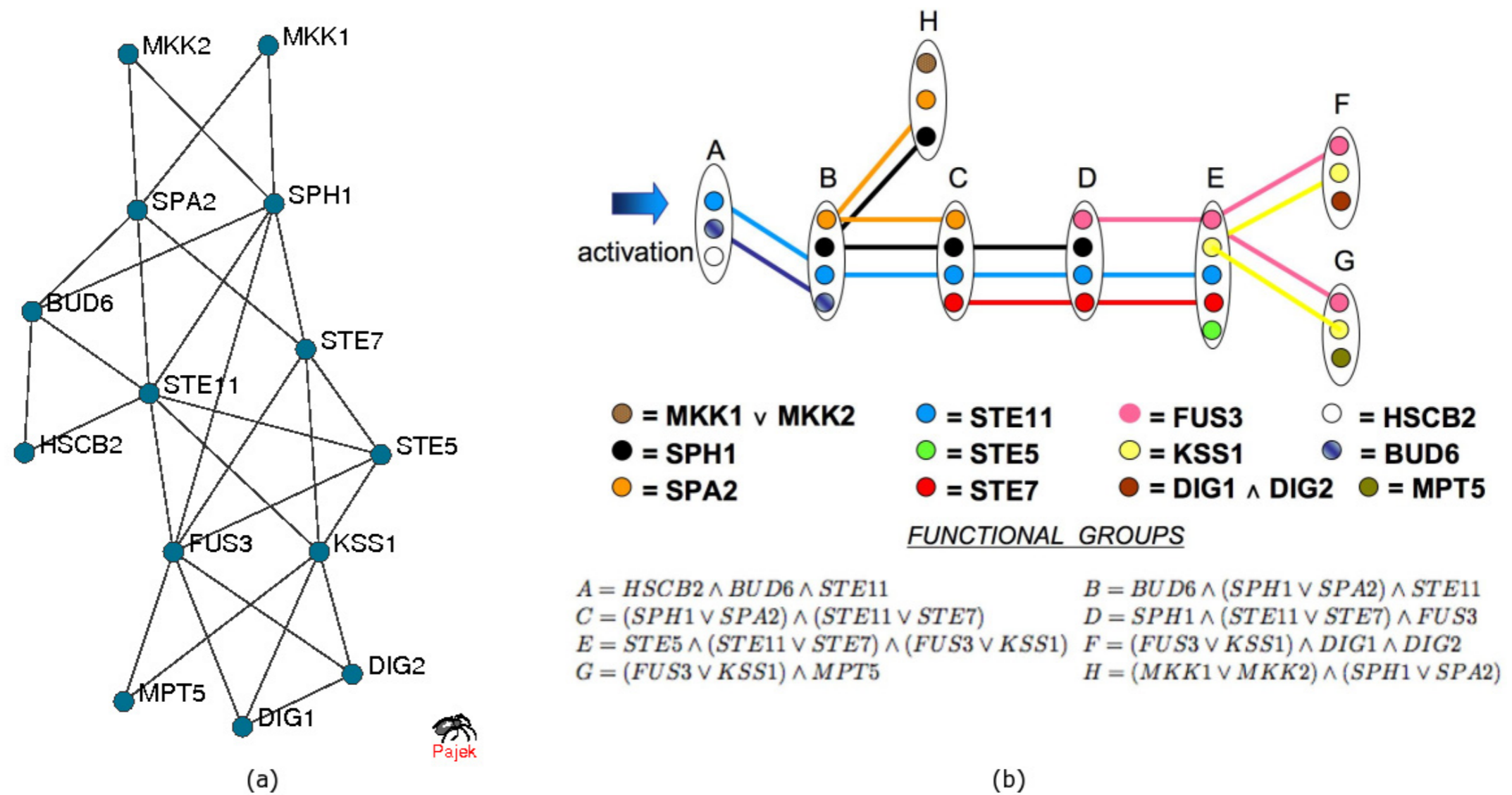
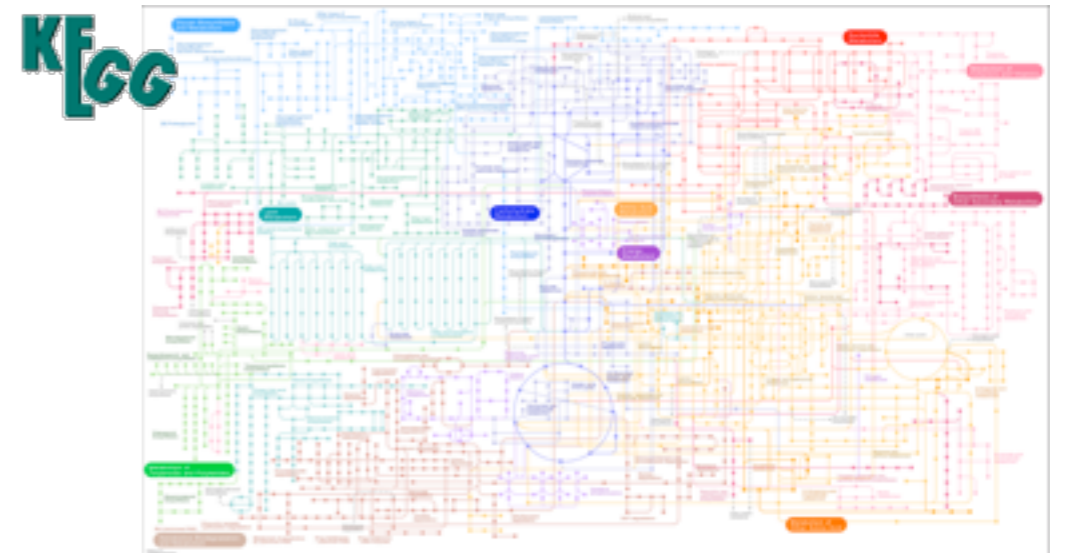


Figure 4
Pheromone Signaling Pathway. The pheromone signaling pathway. (a) The network. (b) The Tree of Complexes representation. For the description of the elements of the tree see the text.

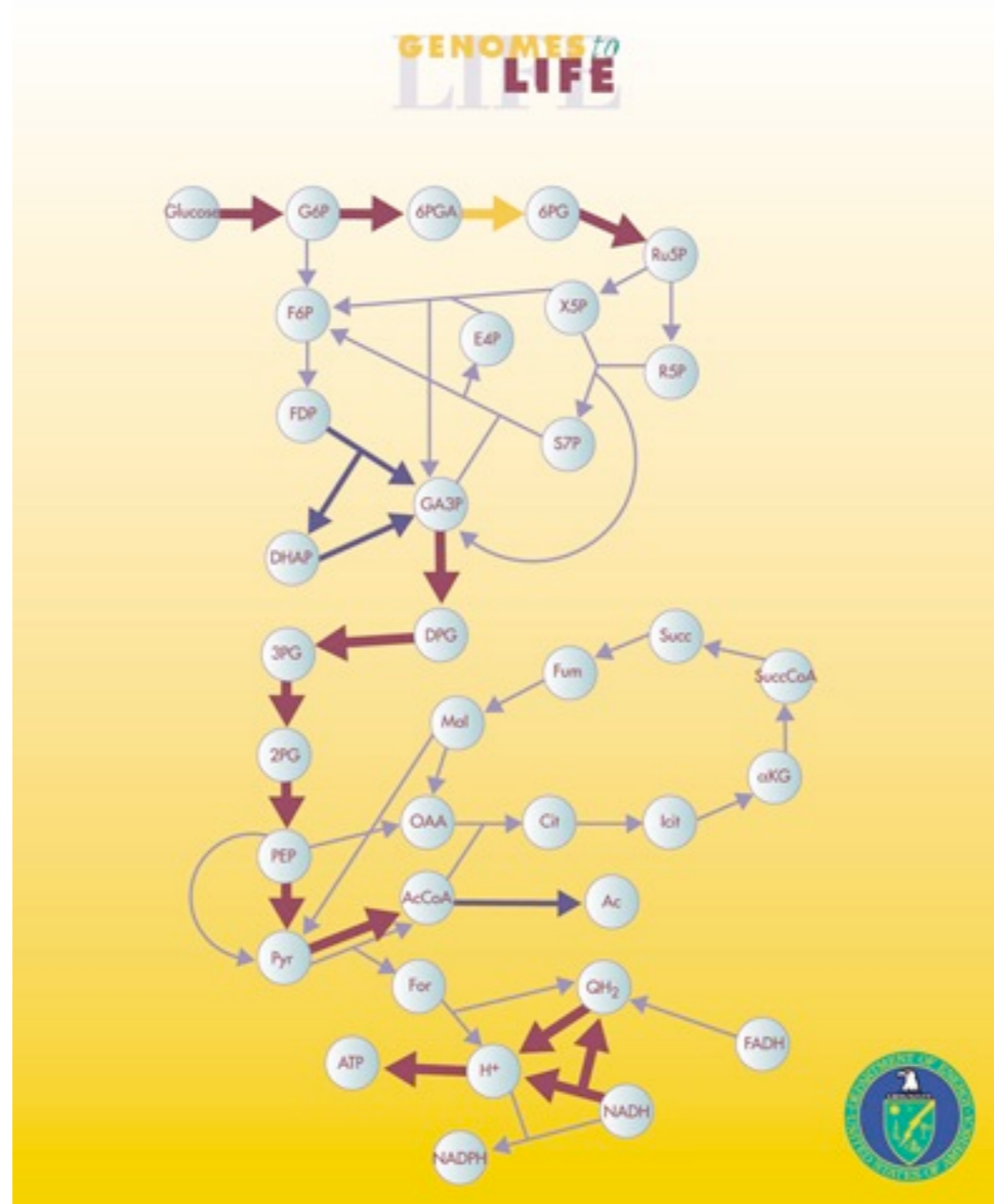
4. Descriptions of Metabolic Networks

- Several databases exist for gene/enzyme/reaction/pathway reconstruction of metabolic pathways
- KEGG, Meta-Cyc, Pathway Tools, ERGO, metaTIGER, ENZYME, Brenda, ModelSEED



Descriptions of Metabolic Networks

- Metabolite-centric views



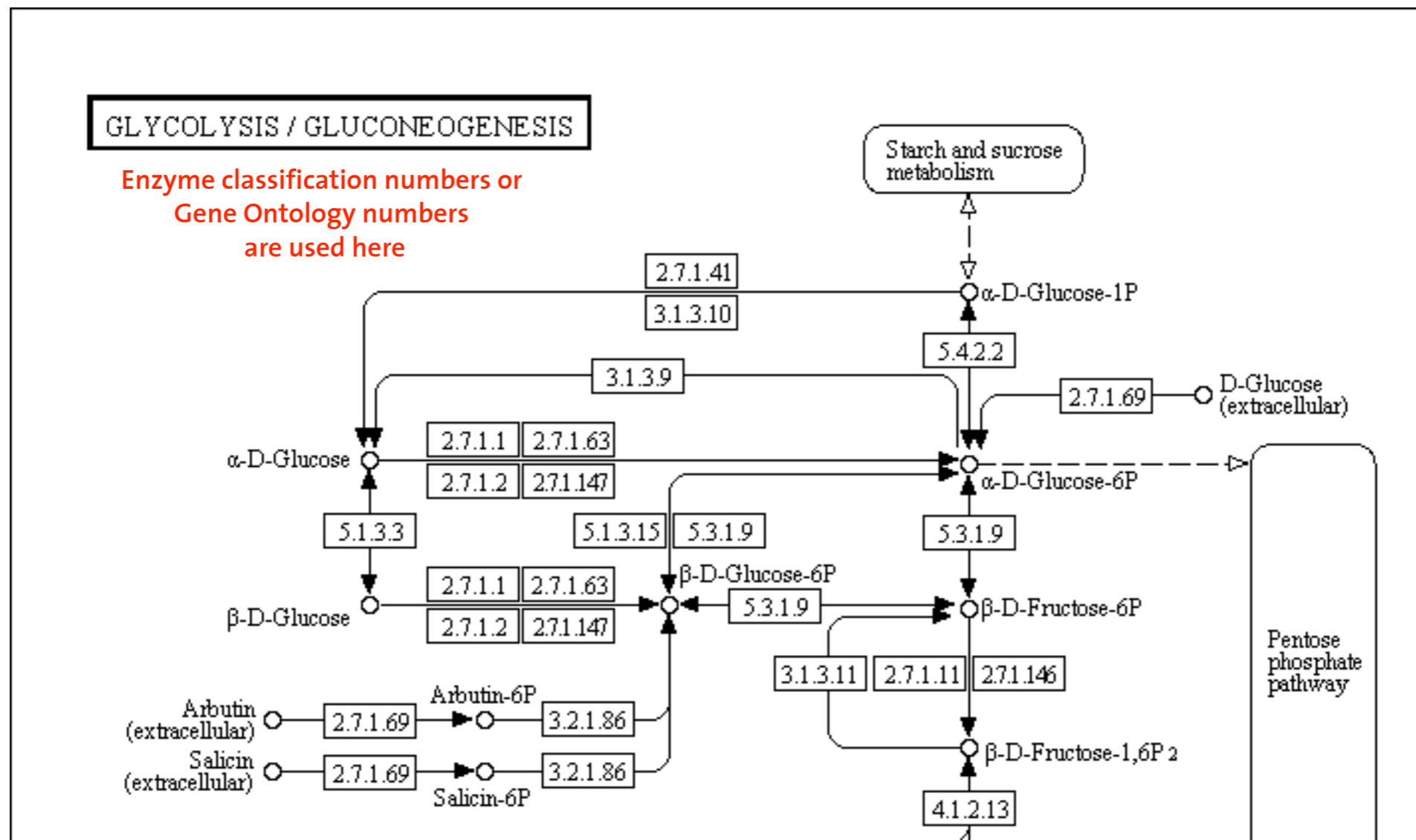
Biotransformations

Nodes are metabolites
Links are enzymes



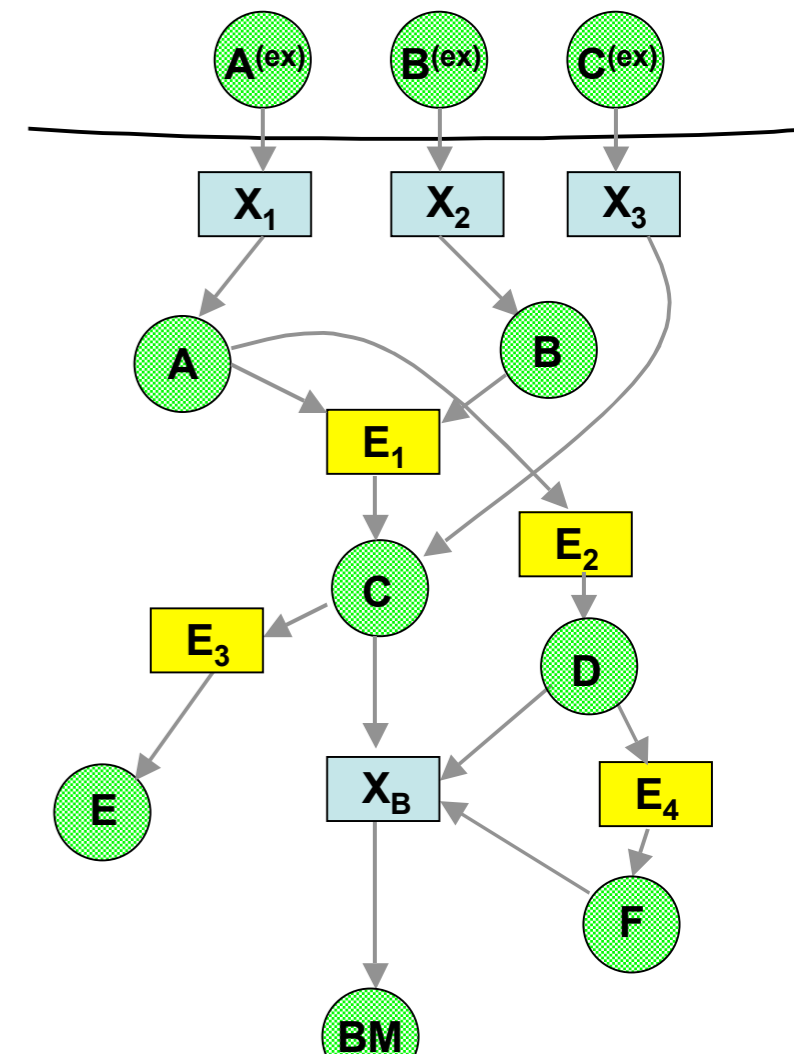
Descriptions of Metabolic Networks

- Metabolite-centric views
- Gene/Enzyme/Reaction-centric views



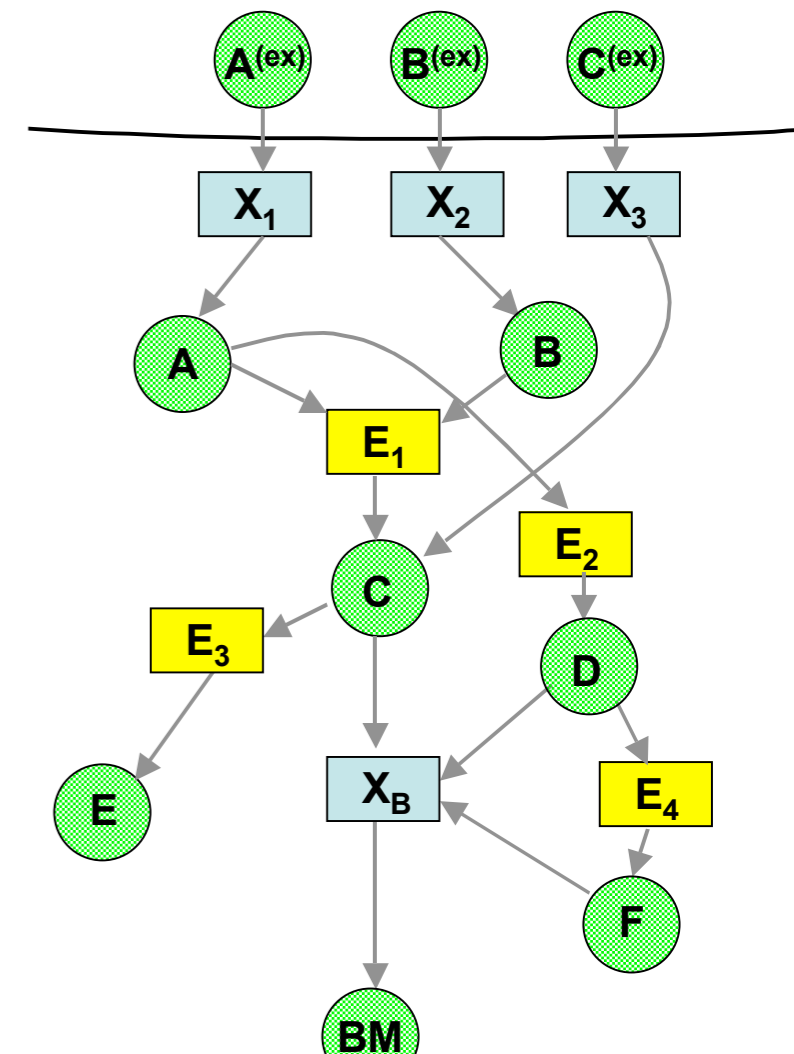
Descriptions of Metabolic Networks

- Metabolite-centric views
- Gene/Enzyme/Reaction-centric views
- Bipartite representation

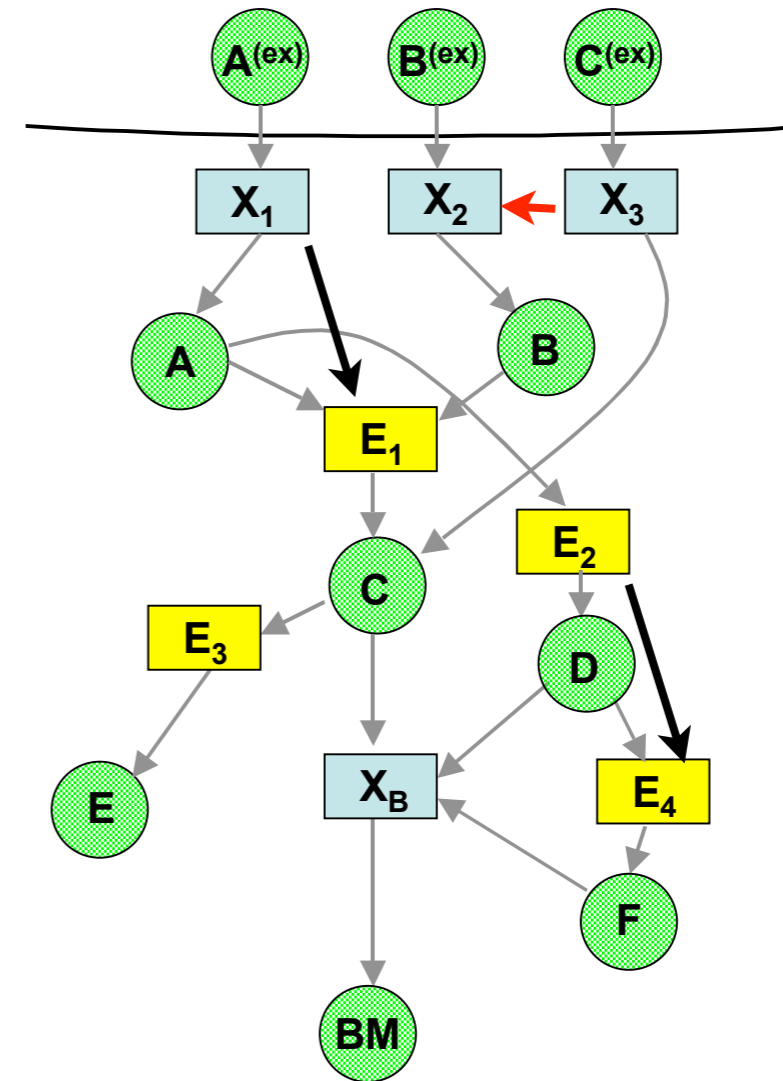
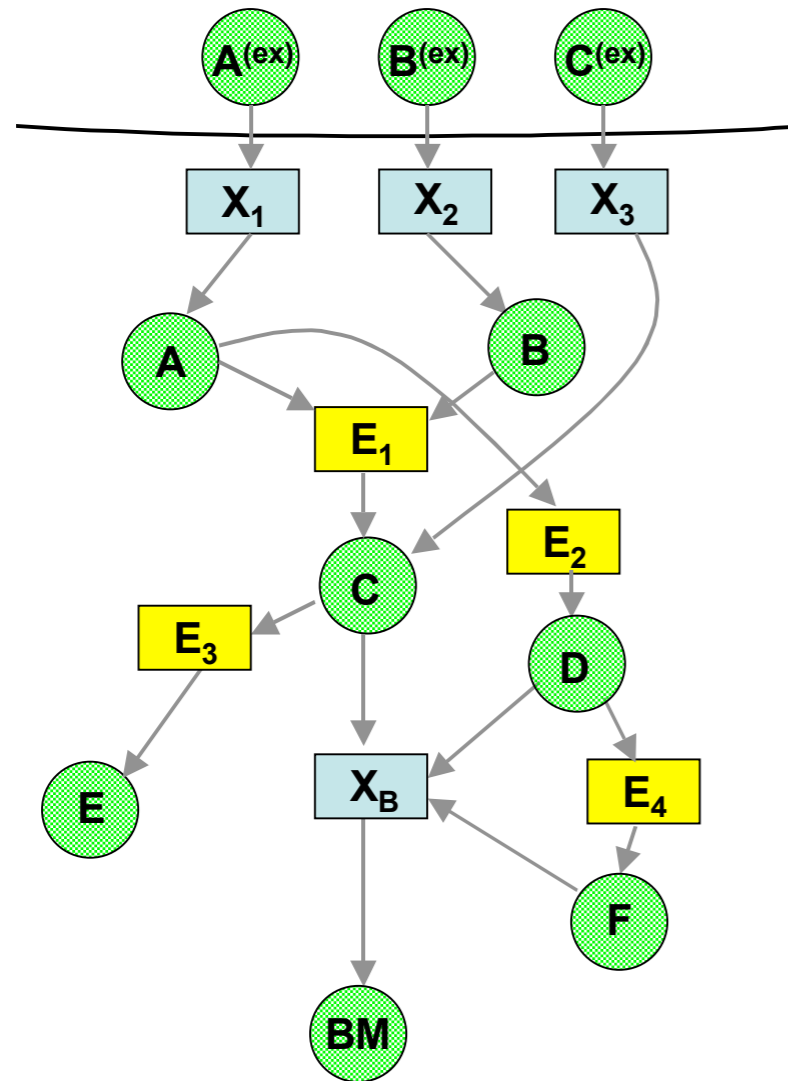


Descriptions of Metabolic Networks



- Metabolite-centric views
- Gene/Enzyme/Reaction-centric views
- Bipartite representation
- Could we have a multipartite representation for PPIs?



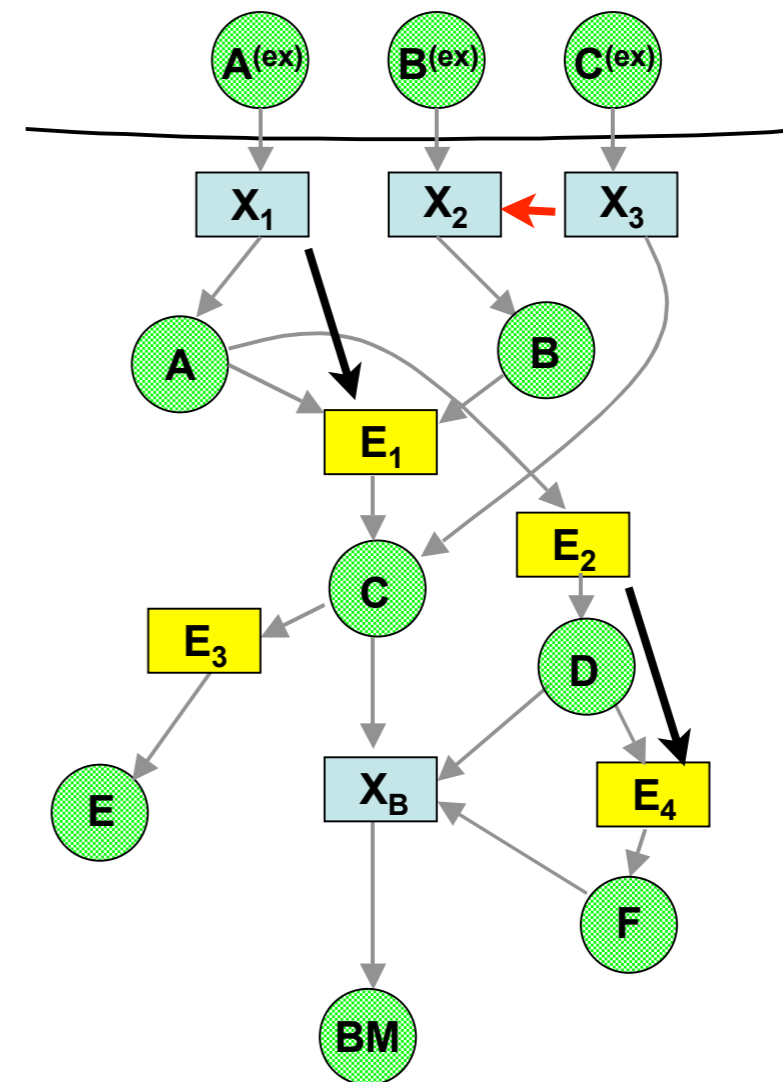
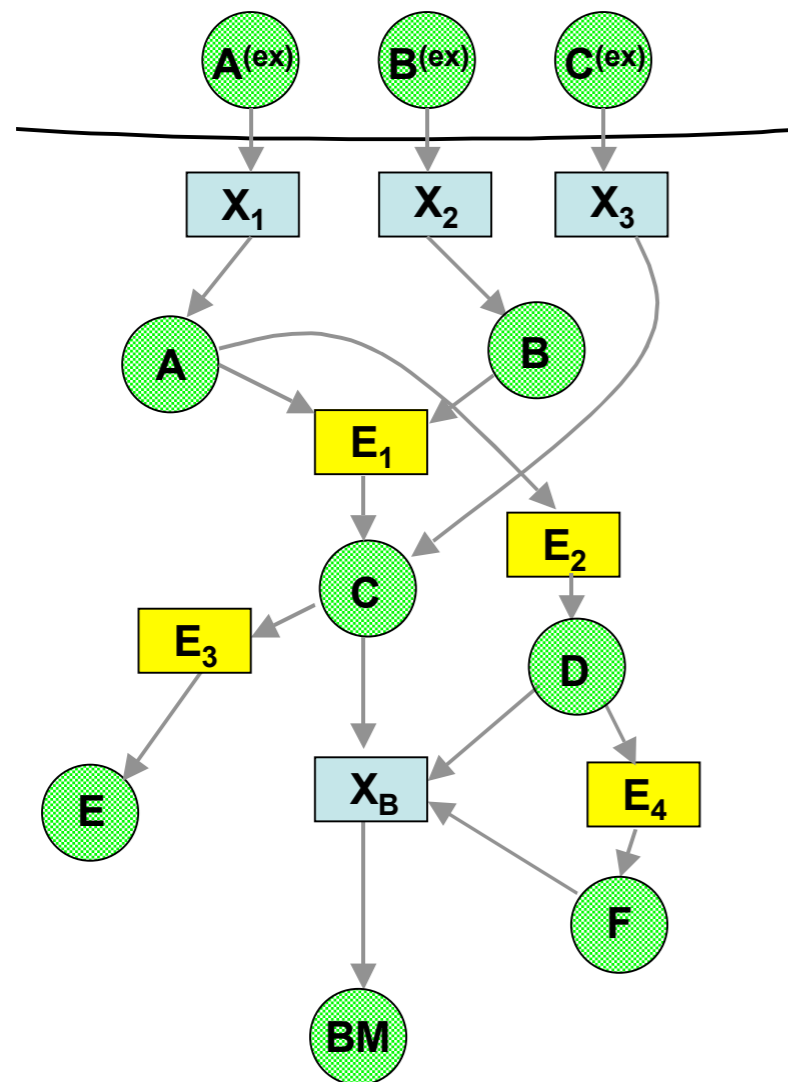
Example Multipartite Representation



**Addition of PPIs to the network
(edges define physical associations)**

-  direction of (obligatory) transient PPI on cytoskeleton for catalysis of metabolic intermediate
-  direction of (obligatory) permanent PPI in membrane for transport of metabolites

Example Multipartite Representation



- Adding PPIs to the network reveals local functionality
- Adding biophysical information describes the nature of the PPIs

Addition of PPIs to the network
(edges define physical associations)

- direction of (obligatory) transient PPI on cytoskeleton for catalysis of metabolic intermediate
- direction of (obligatory) permanent PPI in membrane for transport of metabolites

Network-based prediction of metabolic enzymes' subcellular localization

Shira Mintz-Oron^{1,*}, Asaph Aharoni¹, Eytan Ruppin^{2,3} and Tomer Shlomi^{4,*}

¹Department of Plant Sciences, Weizmann Institute of Science, Rehovot 76100, ²School of Computer Science,

³School of Medicine, Tel-Aviv University, Tel-Aviv 69978 and ⁴Faculty of Computer Science, Technion, Haifa 32000, Israel

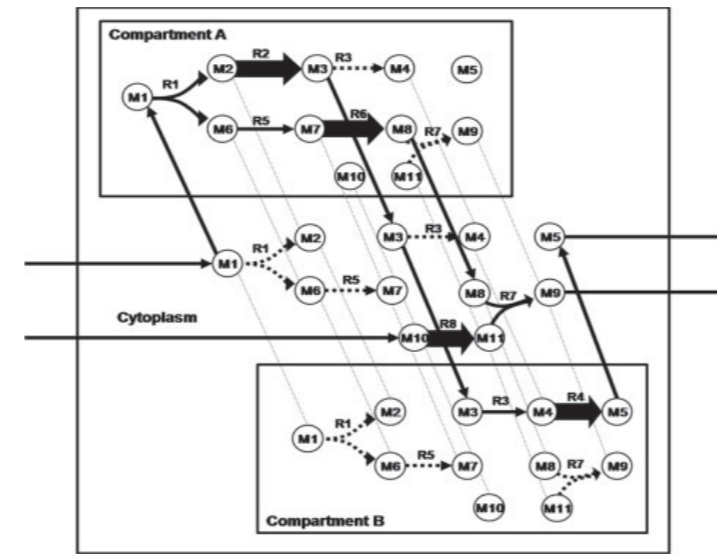
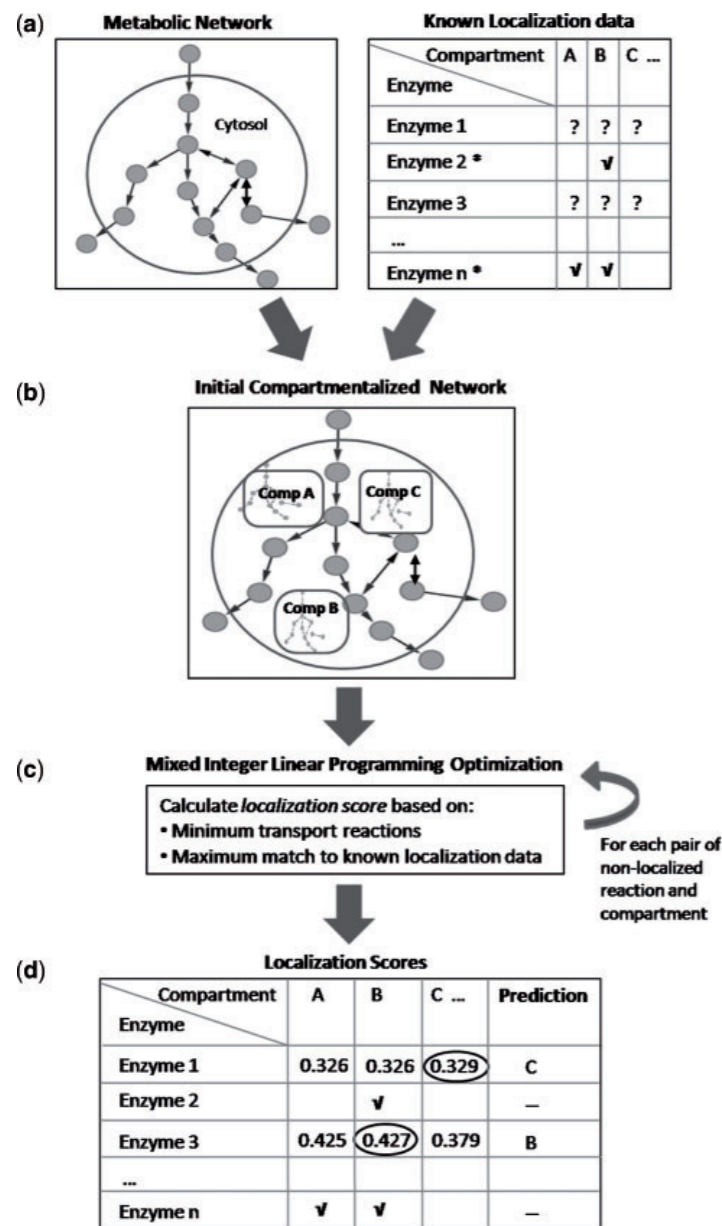


Fig. 2. An illustrative example of our method for enzyme subcellular compartment prediction. The initial compartmentalized network consists of three compartments, with instances of 11 metabolites and 8 reactions in each compartment. Thin edges connecting different instances of a metabolite in various compartments represent transport reactions that move metabolites across membrane boundaries. Wide arrows represent localized reactions whose known localization is given as input to the prediction method. Solid arrows represent reactions that are predicted to have non-zero flux by our method reflecting their predicted localization. Dashed arrows represent reactions predicted to have zero flux.

Fig. 1. A schematic representation of the enzyme subcellular localization prediction method. (a) The input data is a metabolic network, representing a set of enzyme-catalyzed reactions, and the known localization data for a subset of enzymes. (b) Integrating the given network and localization data yields an initial compartmentalized network, consisting of several compartments while the non-localized reactions are duplicated to all compartments. (c) Mixed Integer Linear Programming (MILP) is applied for each pair of non-localized reaction and compartment to calculate a localization score, reflecting the likelihood of this reaction to be present in that compartment. (d) Enzymes are predicted to be localized in compartments achieving the highest localization scores.

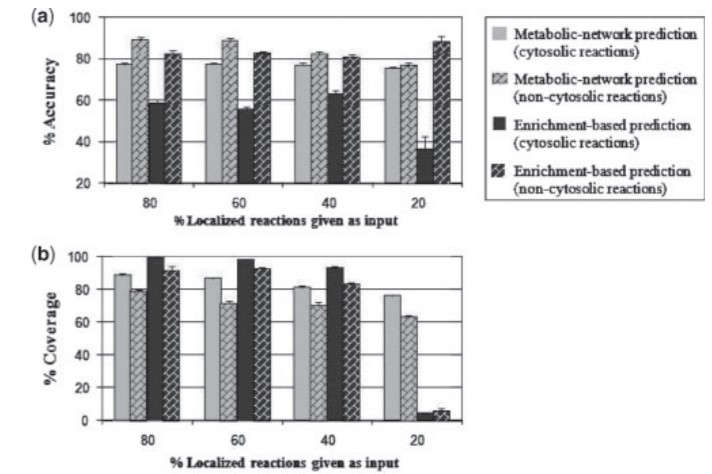


Fig. 3. Accuracy (a) and coverage (b) of enzyme subcellular localization predictions in a cross-validation test in the yeast *S.cerevisiae*. The average and standard error of the accuracy and coverage measures were calculated based on 10 applications of the prediction methods over randomly sampled sets of localized enzymes of similar size that are used as input.

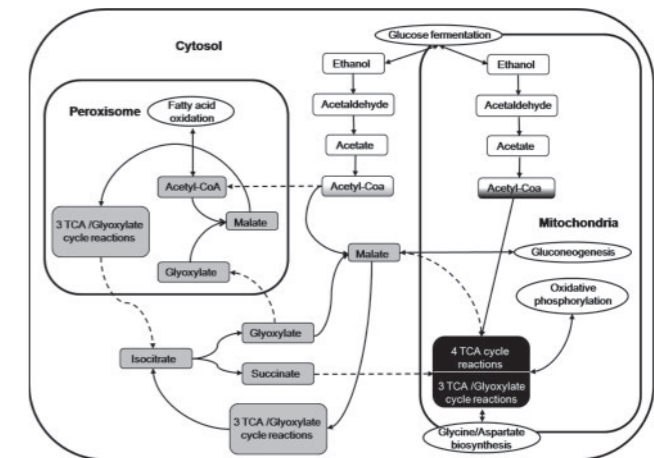


Fig. 4. Enzyme subcellular localization prediction of two complete metabolic pathways, including the TCA cycle (black rectangles) and glyoxylate cycle (grey rectangles), and a subset of the ethanol oxidation pathway (white rectangles), given localization data for enzymes in other connected pathways (white ellipses) as input. Transport reactions are marked by dotted arrows.

5. Functional units of PPIs

Descriptions for functional units of PPIs

- Small scale: as network motifs (directed, weighted, associated with ΔG of binding)



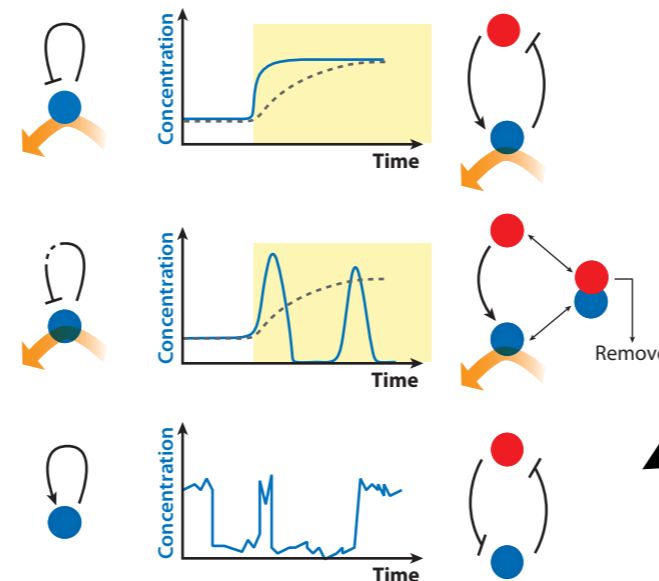
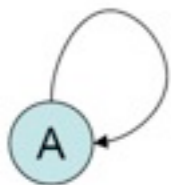
Simplified Models of Biological Networks

Kim Sneppen, Sandeep Krishna, and Szabolcs Semsey

Annu. Rev. Biophys. 2010. 39:43–59

DYNAMICS OF SINGLE FEEDBACK LOOPS

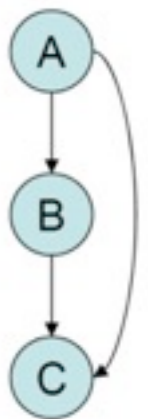
Auto regulatory single protein



Auto regulatory PPIs

Feedback

Feedforward



Schematic illustration of negative and positive feedback loops along with their respective dynamical behavior. (a) Negative feedback stabilizes the output to a near-constant level and allows for fast transient increase in production in response to stress or perturbations. (b) If negative feedback is delayed, the protein concentrations may oscillate in time. (c) Positive feedback can result in bistability, i.e., the system can exist stably in either of two distinct steady states.

Feedback is an essential part of molecular networks. It allows the cell to adjust the repertoire of functional proteins to current needs.

Combinations of FLs in small-molecule uptake and metabolism can result in new behavioral features that are significantly different from a simple sum of the behaviors of single loops.

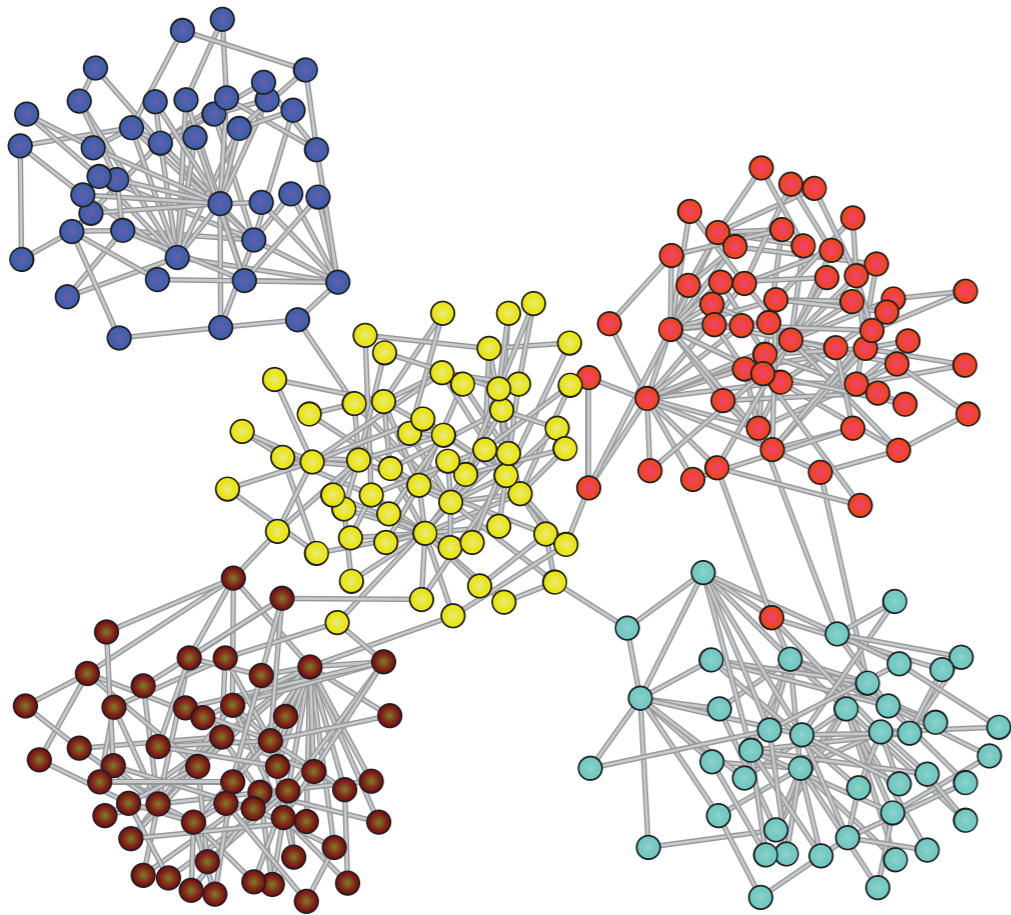
Descriptions for functional units of PPIs

- Small scale: as network motifs (directed, weighted, associated with ΔG of binding)
- Larger scale: Location of PPI classes - transient and permanent, obligate & non-obligate
- Larger scale: Classes of PPIs associated with “Hubs” and “Loners” and essentiality.

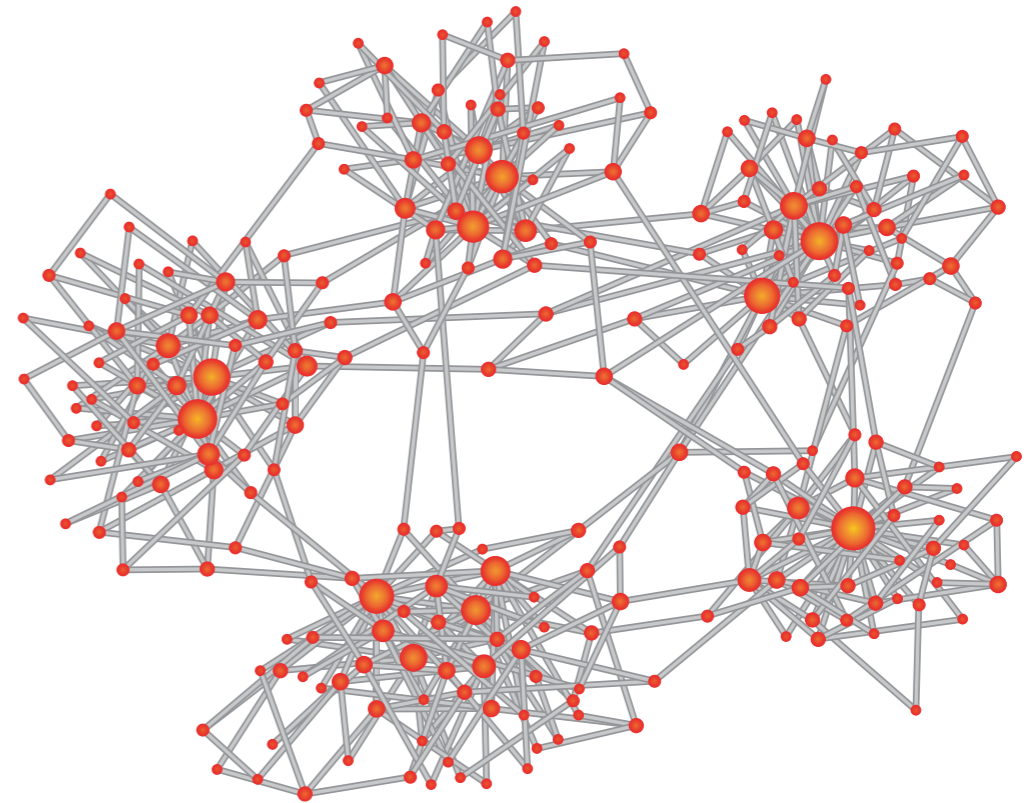
Defining PPIs essential to metabolism

- PPIs associated network motifs as “hubs” in modules would be critical features in the metabolic network
- PPIs associated network motifs as inter-module links would also be critical to a metabolic network

Close large scale models of metabolic pathways



Graph of a modular scale-free structure



Graph of distributed “hubs” in modules

Integrated Bioinformatics for Radiation-Induced Pathway Analysis from Proteomics and Microarray Data

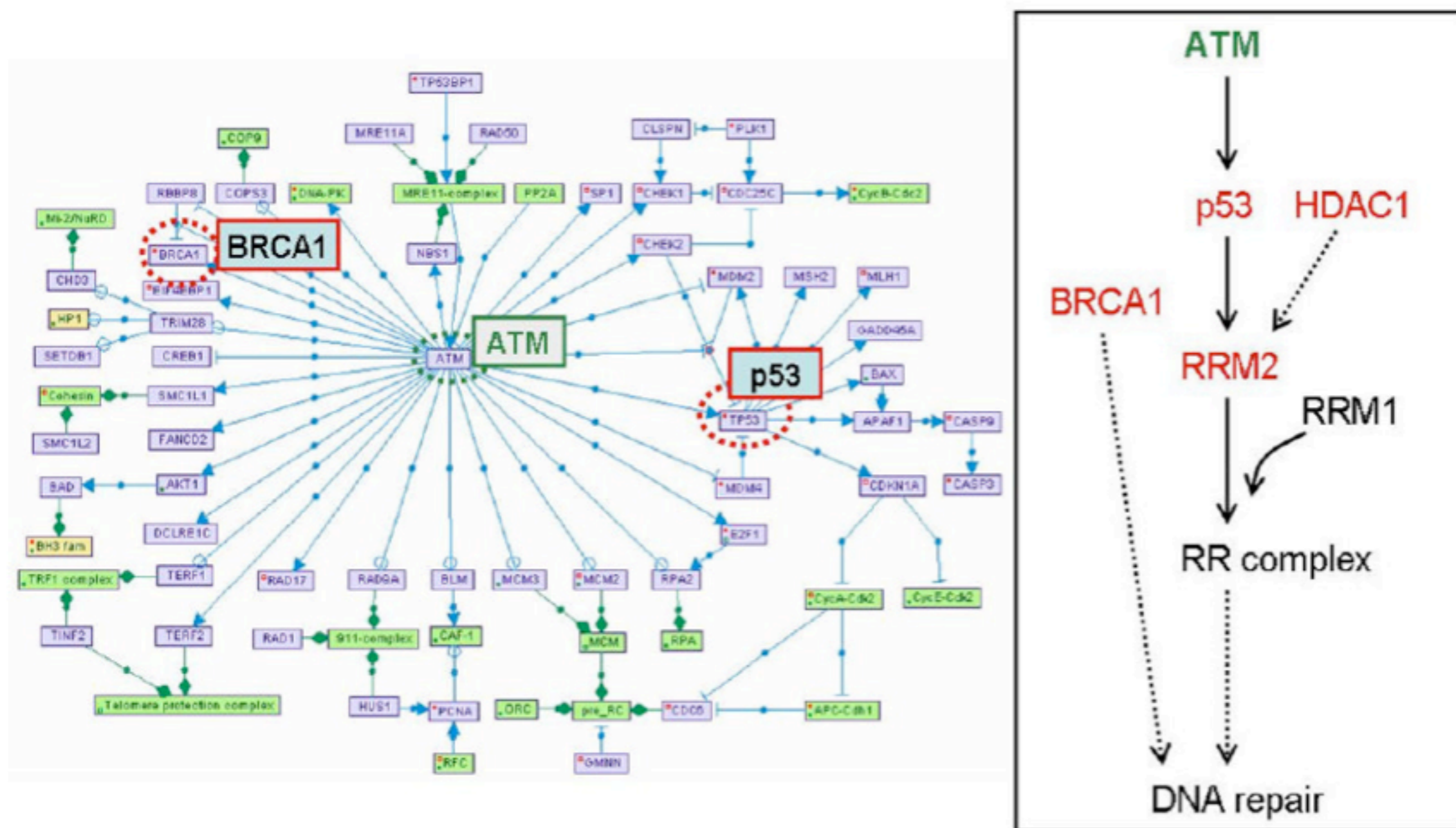


Figure 6. RRM2 is involved in radiation-induced ATM-p53-mediated DNA repair pathway

Example References

Biochemical Society Transactions (2008) Volume 36, part 6

Topological properties of protein interaction networks from a structural perspective

Attila Gursoy*, Ozlem Keskin* and Ruth Nussinov†‡¹

*Center for Computational Biology and Bioinformatics and College of Engineering, Koc University, Rumelifeneri Yolu, 34450 Sariyer Istanbul, Turkey, †Basic Research Program, SAIC-Frederick, Inc., Center for Cancer Research Nanobiology Program, NCI-Frederick, Frederick, MD 21702, U.S.A., and ‡Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

[IEEE/ACM Trans Comput Biol Bioinform.](#) 2011 Nov 10. [Epub ahead of print]

Identification of Essential Proteins Based on Edge Clustering Coefficient.

[Wang J](#), [Li M](#), [Wang H](#), [Pan Y](#).

Central South University, ChangSha.

In silico network topology-based prediction of gene essentiality

João Paulo Müller da Silva, Marcio Luis Acencio, and Ney Lemke
*Department of Physics and Biophysics, Institute of Biosciences,
São Paulo State University, UNESP, 18618-000, Botucatu, SP, Brazil*

José Carlos Merino Mombach
*Centro de Ciências Rurais, Unipampa/São Gabriel - Pós-Graduação em Física,
Prédio 13, Universidade Federal de Santa Maria, Santa Maria 97105-900, Brazil*

Renata Vieira, José Camargo da Silva, and Marialva Sinigaglia
*Laboratório de Bioinformática e Biologia Computacional Programa Interdisciplinar de Computação Aplicada,
Universidade do Vale do Rio dos Sinos, 93022-000 São Leopoldo, RS, Brazil*

(Dated: February 2, 2008)

Physica A: Statistical Mechanics and its Applications (2007)

Volume: 387, Issue: 4, Pages: 1049-1055

arXiv: [0709.4206](#)

Summary

- **Structural and experimental data reveals a wealth of PPIs**
- **Biophysical information defines the nature and class of PPIs**
- **PPIs are poorly represented in network descriptions**
- **Incorporating PPI biophysics could better describe the distribution of functional units in Metabolic networks.**