



MASTER IN TECNOLOGIE BIOINFORMATICHE APPLICATE ALLA  
MEDICINA PERSONALIZZATA

---

## Protein Sequence Analysis

# Prediction of Protein-protein Interactions from Evolutionary information

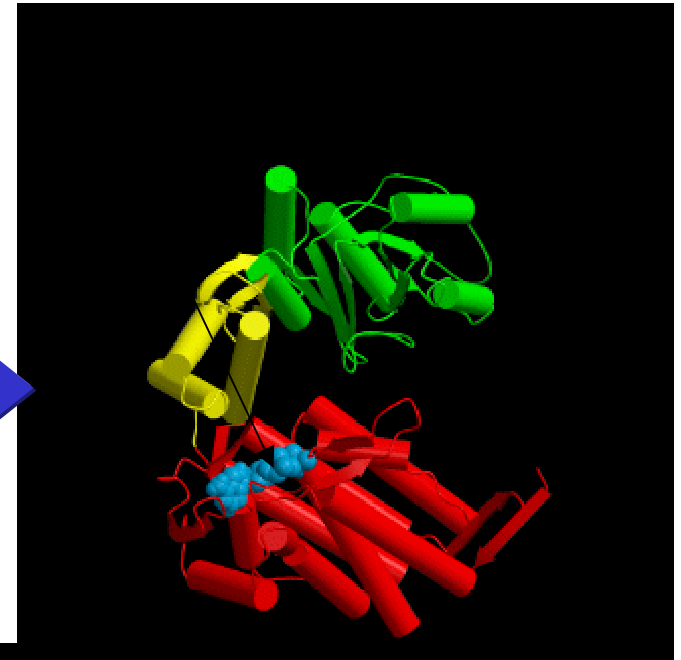
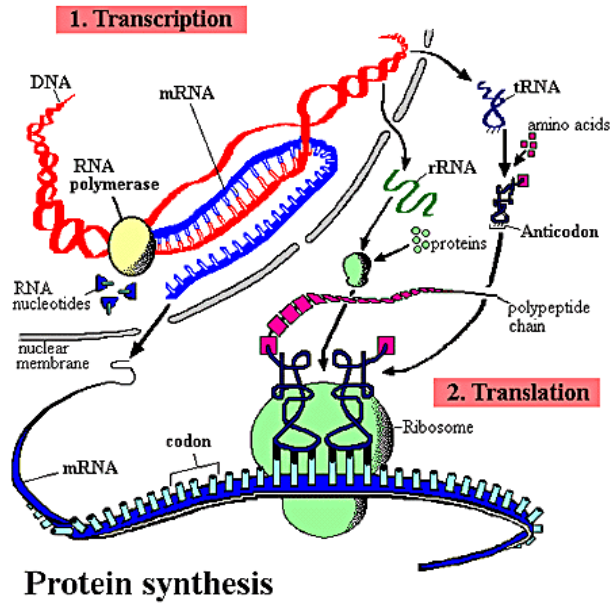
Florencio Pazos (CNB-CSIC)

---

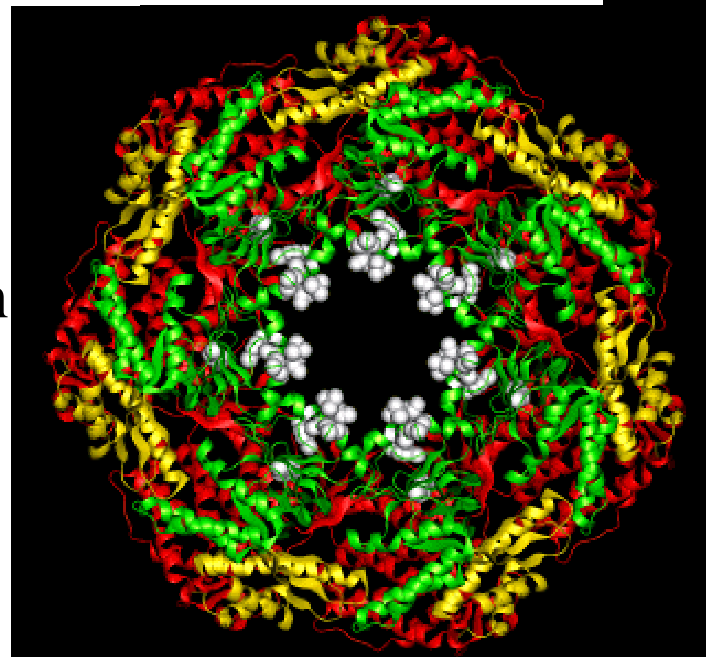
*Florencio Pazos Cabaleiro*  
*Protein Design Group (CNB-CSIC)*  
*pazos@cnb.uam.es*



# Protein sequences, structures and functions



subunit

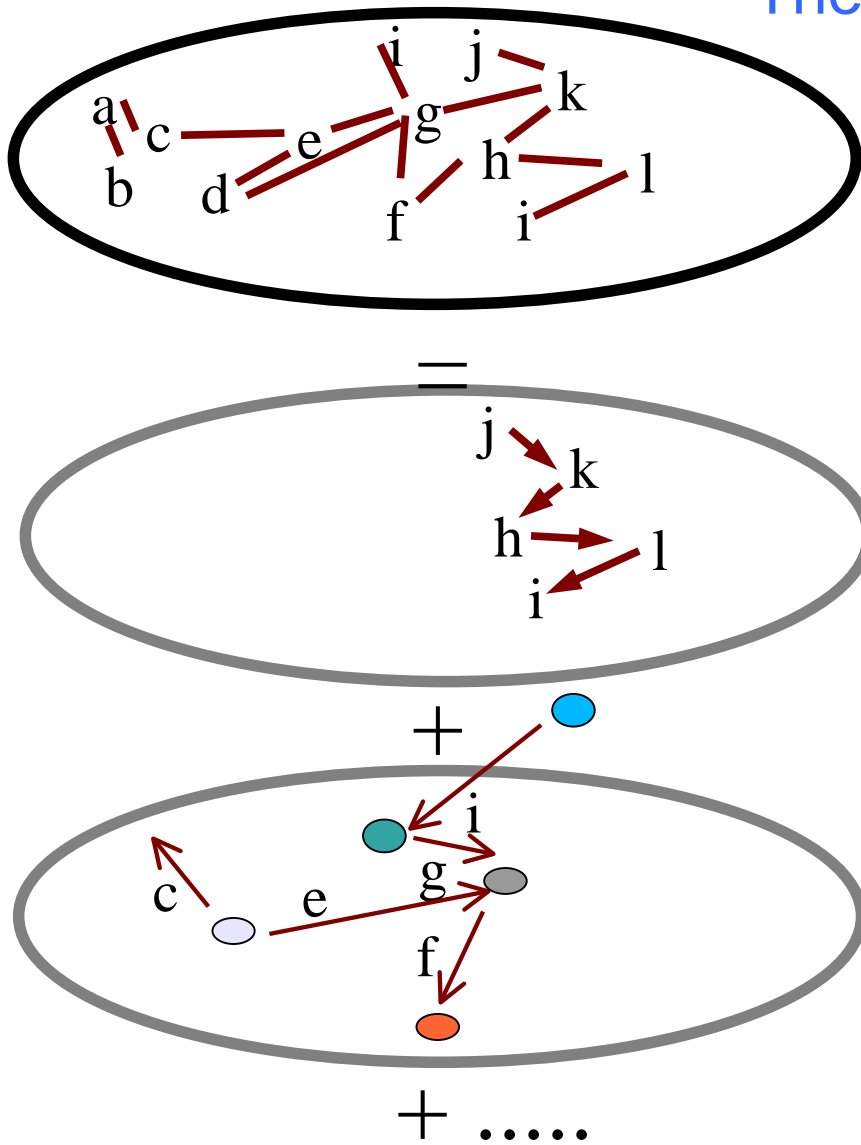


heptamer

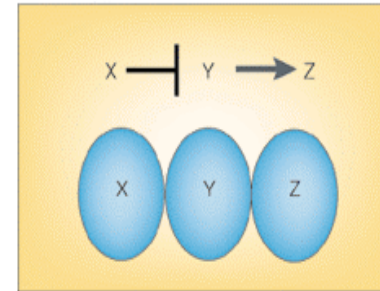
Molecular chaperonin  
GroEL

(Dr Jianpeng Ma, Harvard Univ.)

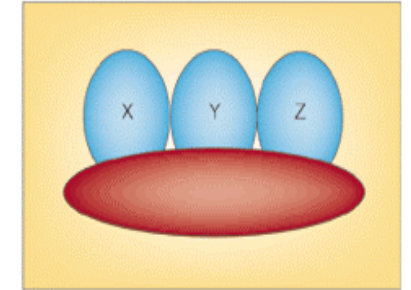
# The interactome



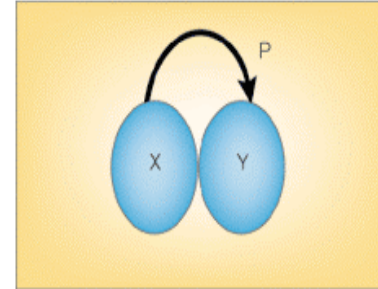
**a** Genetic pathways



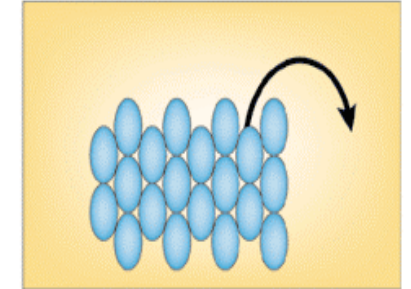
**b** Pathway scaffolding



**c** Enzymatic reactions



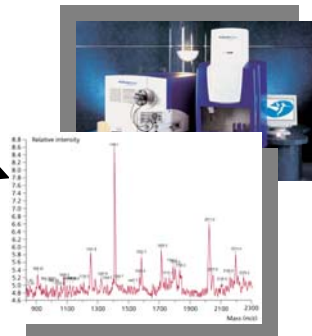
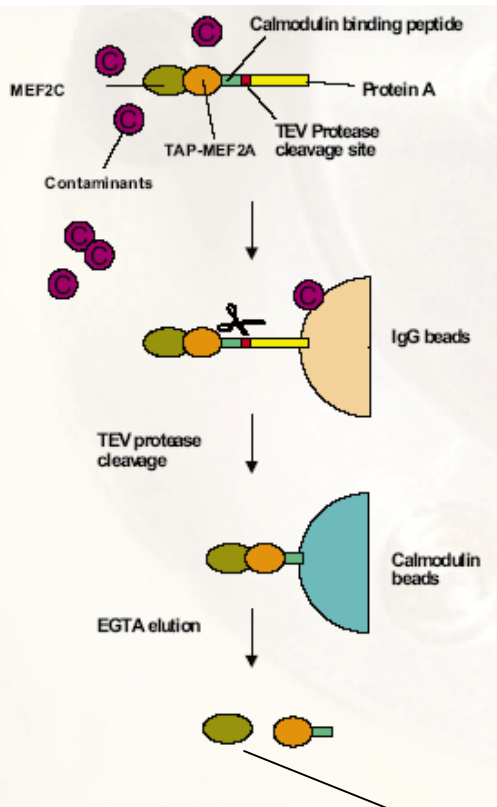
**d** Molecular machines



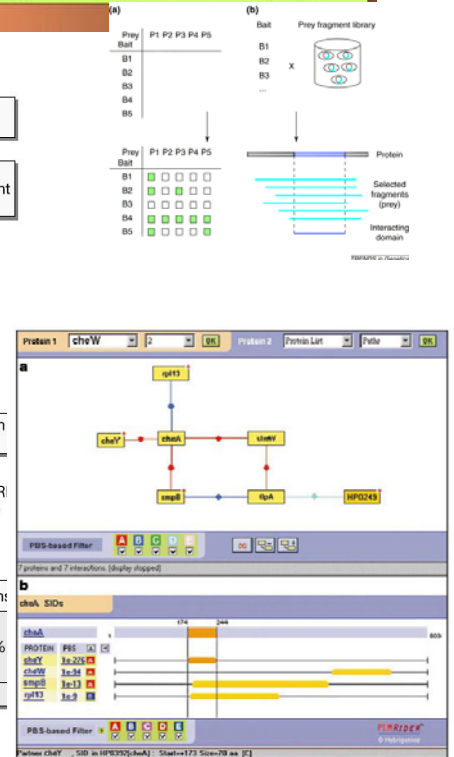
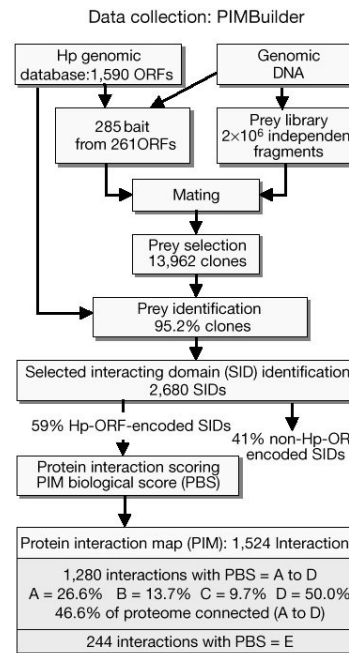
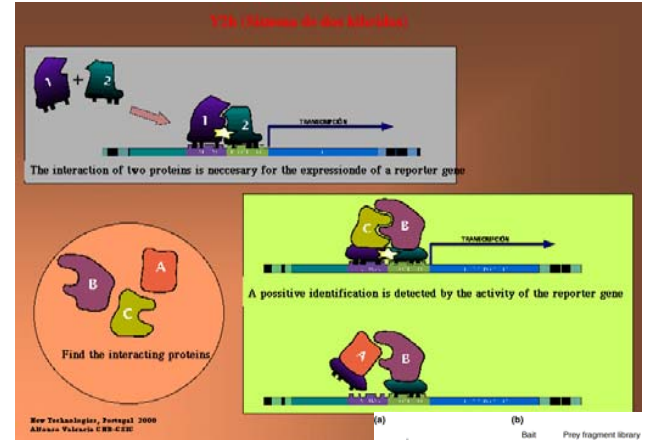
Nature Reviews | Molecular Cell Biology

# Experimental determination of the interactome

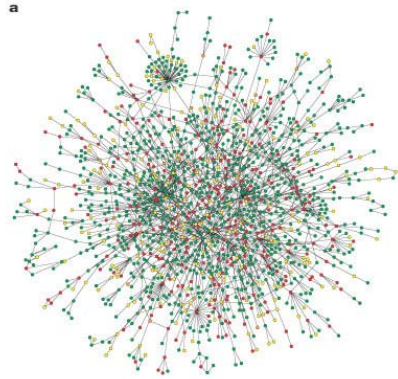
## TAP/MS



## Y2H



# Protein Interaction Networks (“interactome”)



- 
- Rain, J.C., Selig, L., De Reuse, H., et al. (2001) The protein-protein interaction map of *Helicobacter pylori*. *Nature*, **409**, 211-215.
  - Gavin, A.C., et al. (2002) Functional organisation of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**, 141-147.
  - Ho, Y., et al. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, **415**, 180-183.
  - Ito, T., et al. (2000) Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc Natl Acad Sci USA*, **97**, 1143-1147.
  - Uetz, P., et al. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*, **403**, 623-631.
  - Giot, L., Bader, J.S., Brouwer, et al. (2003) A protein interaction map of *Drosophila melanogaster*. *Science*, **302**, 1727-1736.
  - Li, S., Armstrong, C.M., Bertin, N., et al. (2004) A map of the interactome network of the metazoan *C. elegans*. *Science*, **303**, 540-543.
  - Butland, G., Peregrin-Alvarez, J.M., Li, J., et al. (2005) Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature*, **433**, 531-537.
  - Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., et al. (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, **122**, 957-968.
  - Uetz, P., Dong, Y.A., Zeretzke, C., Atzler, C., Baiker, A., Berger, B., Rajagopala, S.V., Roupelieva, M., Rose, D., Fossum, E., et al. (2006) Herpesviral protein networks and their interaction with the human proteome. *Science*, **311**, 239-242.
  - LaCount, D.J., Vignali, M., Chettier, R., Phansalkar, A., Bell, R., Hesselberth, J.R., Schoenfeld, L.W., Ota, I., Sahasrabudhe, S., Kurschner, C., et al. (2005) A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature*, **438**, 103-107.
  - Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., et al. (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, **437**, 1173-1178.
  - Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., et al. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature*, **440**, 631-636.
  - Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature*, **440**, 637-643.

# Quality of the high-throughput interaction data

Table 1  
Large protein interaction screens for eukaryotes

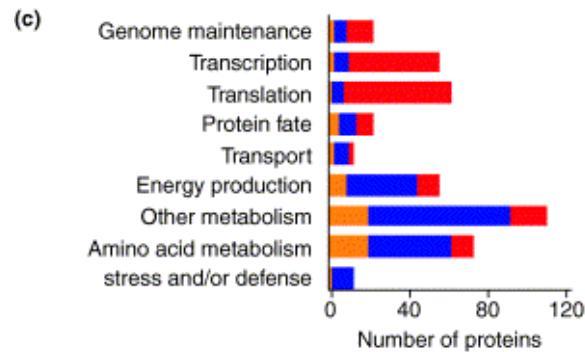
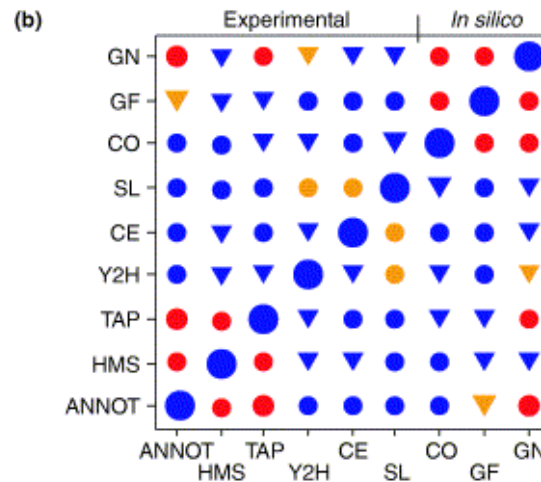
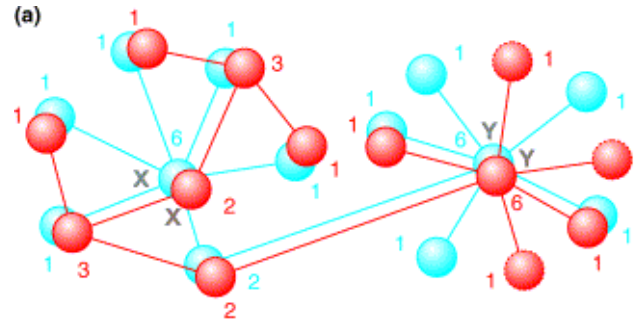
Organism (genes)	Method	Interactions <sup>a</sup>	Proteins	Reference
Yeast (~6000)	Yeast two-hybrid	967	1004	[63]
	Yeast two-hybrid	4549	3278	[13]
	Yeast two-hybrid	420	271	[64,65]
	Co-AP/MS	9421	1665	[66]
	Co-AP/MS	3878	1578	[67]
<i>Drosophila</i> (~14 000)	Yeast two-hybrid	20 405	7048	[49]
	Yeast two-hybrid	1814	488	[14]
Worm (~20 000)	Yeast two-hybrid	4027	1926	[68]

<sup>a</sup>For two-hybrid screens, the approximate number of unique binary interactions is shown. For co-AP/MS screens, the approximate number of binary interactions that would result if each bait protein contacted every protein that co-purified with it (the “hub and spoke” model) is shown. Data can be retrieved from one of the databases cited [42–44].

Overlap:  
6 int !

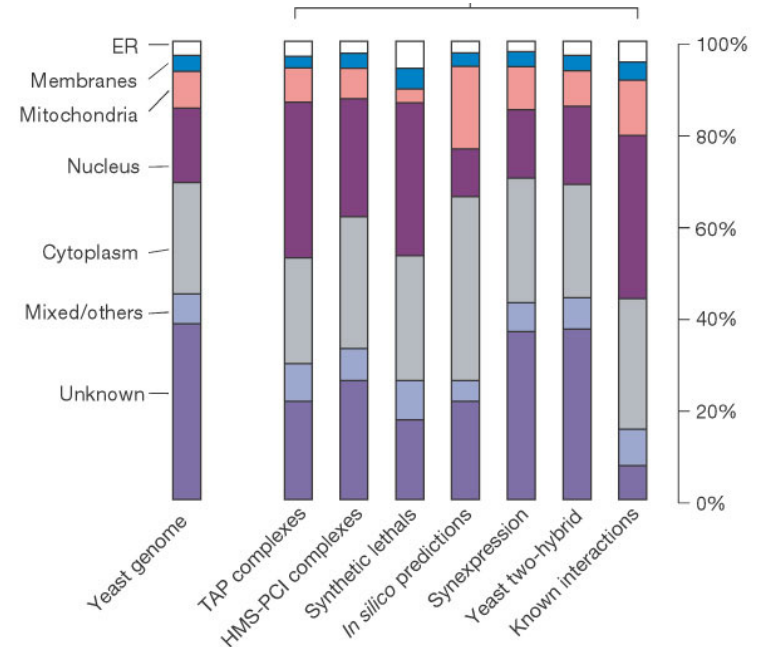
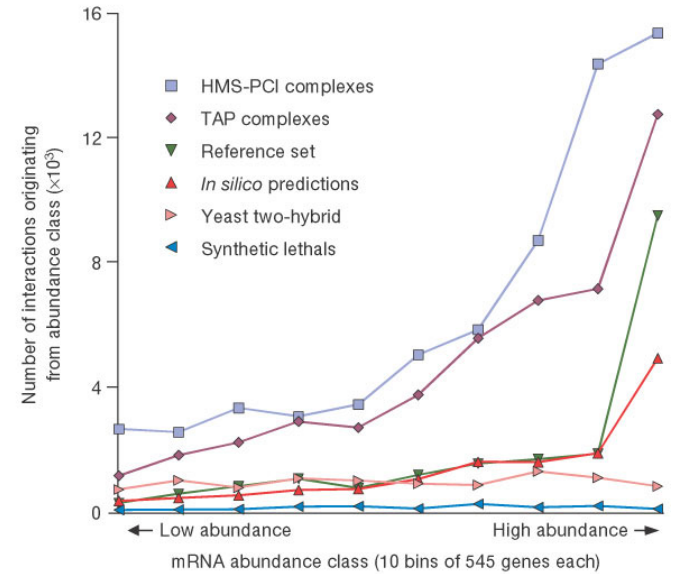
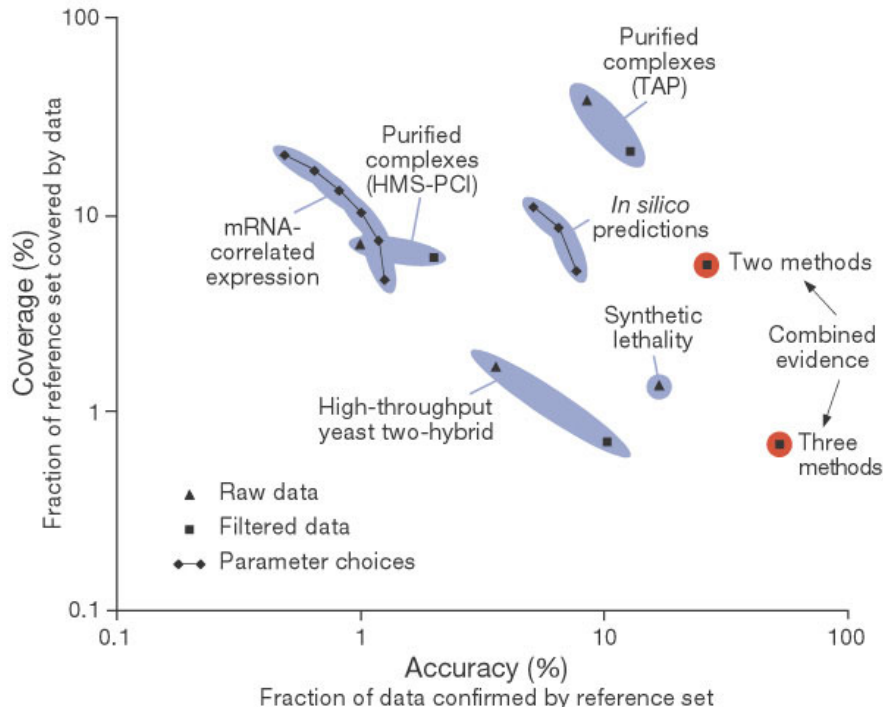
Estimation (yeast): 12.000-40000 ints. (6000 prots.)

# Quality of the high-throughput interaction data





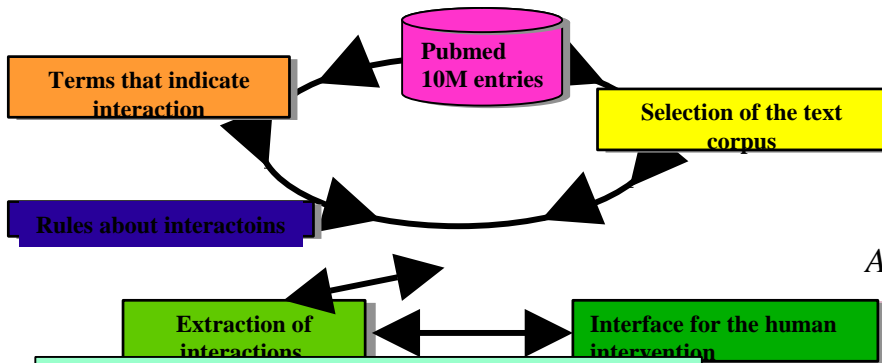
# Quality of the high-throughput interaction data



von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S. and Bork, P. (2002) Comparative assessment of large scale data sets of protein-protein interactions. *Nature*, **417**, 399-403.



# Retrieving protein relationships from the literature (text mining)

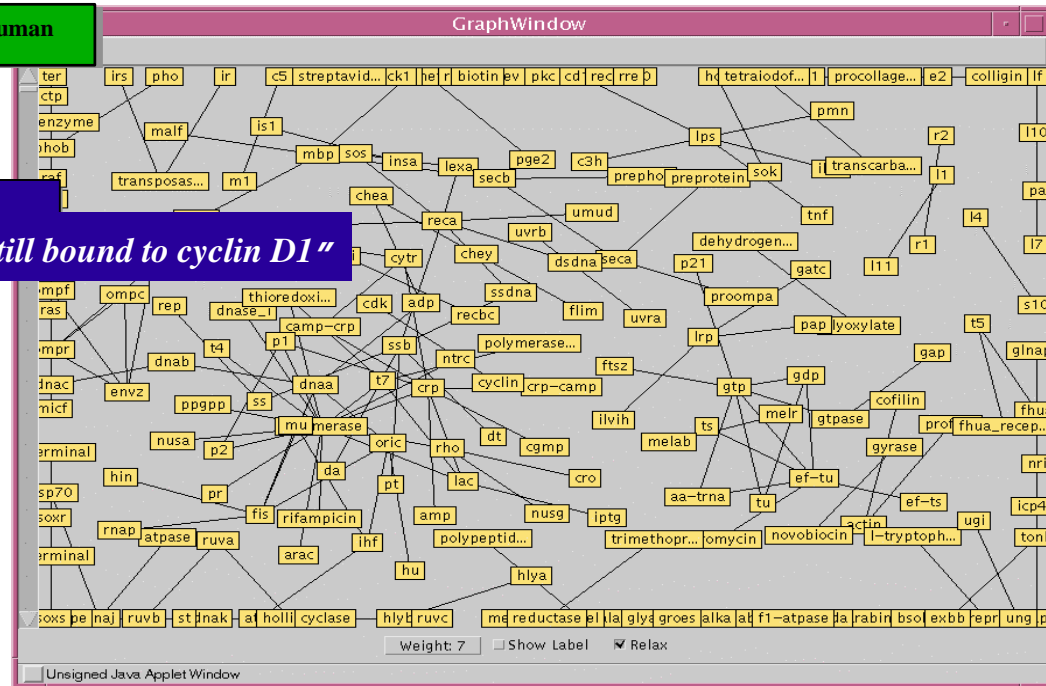


A. Valencia

Action words are for example:

~~bind, associate with, interact, phosphorylate, regulate~~

*"After extensive purification, Cdk2 was still bound to cyclin D1"*



# Retrieving protein relationships from the literature. *iHop*

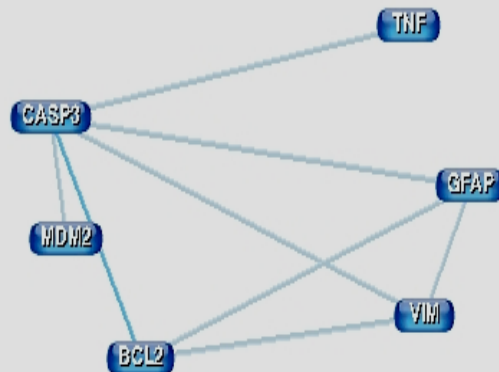
**iHOP**  
 information hyperlinked  
 Over Proteins

Search Gene

Clear model

Print version

Edges in the graph  
 correspond to genes  
 that concur in at least  
 1 2 3 phrases.



MDM2 is **cleaved** by Caspase 3 (CASP3) during apoptosis after aspartic acid-361, generating a 60 kd fragment.

These findings indicate that IR-induced apoptosis involves activation of **CASP3** and that this CrmA-insensitive apoptotic pathway is distinct from those **induced** by **TNF** and certain other stimuli.

At the end of the experiment, the apoptotic markers (BAX, **BCL2** and

Fas-induced activation of the ce

**iHOP**  
 information hyperlinked  
 Over proteins

Search Gene

Show associations of  
 SNF1 with genes  
 from...

Human

Mouse

Drosophila

Zebrafish

C. elegans

Arabidopsis

S. Cerevisiae

E. Coli

Filter and options

Gene Model

Print version

Help

Symbol	Name	Synonyms	Organism
SNF1		CAT1, CCR1, GLC2, HAF3, PAS14	Saccharomyces cerevisiae

NCBI Protein NP\_010765

The **snf1** mutation also **suppresses** the glucose repression defects of **reg1**.

The **SIP1** protein **co-immunoprecipitated** with **SNF1** and was phosphorylated in vitro.

Here we show that **Reg1** **interacts** with the **Snf1** catalytic domain in the two-hybrid system.

Previous studies showed that **Reg1** **regulates** the **Snf1** protein kinase in response to glucose.

The **SNF4** protein is physically **associated** with **SNF1** and positively affects the kinase activity.

The **Sip1** protein is known to undergo phosphorylation when **associated** in vitro with the **Snf1** protein kinase.

Genetic evidence indicated that the catalytic activity of **Snf1** negatively **regulates** its interaction with **Reg1**.

The **SNF1** protein kinase and the **associated SNF4** protein are required for release of glucose repression in Saccharomyces cerevisiae.

The **SIP1** gene of Saccharomyces cerevisiae is a carbon-catabolite-specific negative regulator of GAL gene transcription and acts as a multicopy suppressor of growth defects **associated** with impaired **Snf1p** protein kinase activity.

We show that different sequences of **Reg1** **interact** with **Glc7** and **Snf1**.

In two-hybrid assays, one **SNF4** mutation **enhances** the interaction between **Snf4** and **Snf1**.

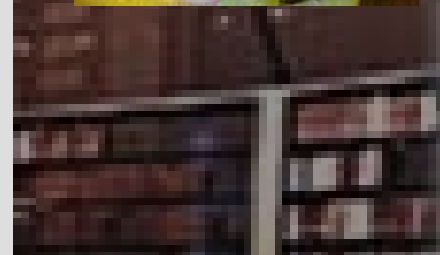
Previously, we identified **SIP1** and **SIP2** as proteins that **interact** with **SNF1** in vivo by the two-hybrid system.

Previous experimental evidence had indicated that **Reg1** might **target** **Glc7** to nuclear substrates such as the **Snf1** kinase complex.

The catalytic subunits of Arabidopsis SnRKs, AKIN10 and AKIN11, interact with **Snf4** and **suppress** the **snf1** and **snf4** mutations in yeast.

**Pak1** **associates** with the **Snf1** kinase in vivo, and the association is greatly enhanced under glucose-limiting conditions when **Snf1** is active.

We show that **SNF4** **binds** to the **SNF1** regulatory domain in low glucose, whereas in high glucose the regulatory domain binds to the kinase domain of **SNF1** itself.



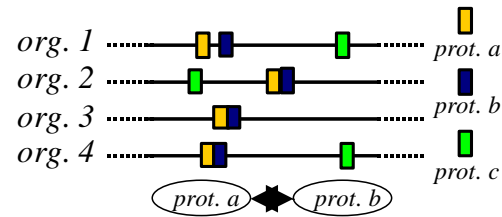
# In-silico prediction of protein interactions

## a) phylogenetic profiles

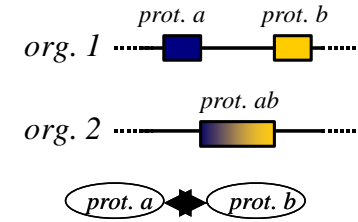
	<i>prot. a</i>	<i>prot. b</i>	<i>prot. c</i>	<i>prot. d</i>
<i>org. 1</i>	1	1	1	1
<i>org. 2</i>	0	1	0	1
<i>org. 3</i>	1	0	1	0
<i>org. 4</i>	1	0	1	1

*prot. a*
★
*prot. c*

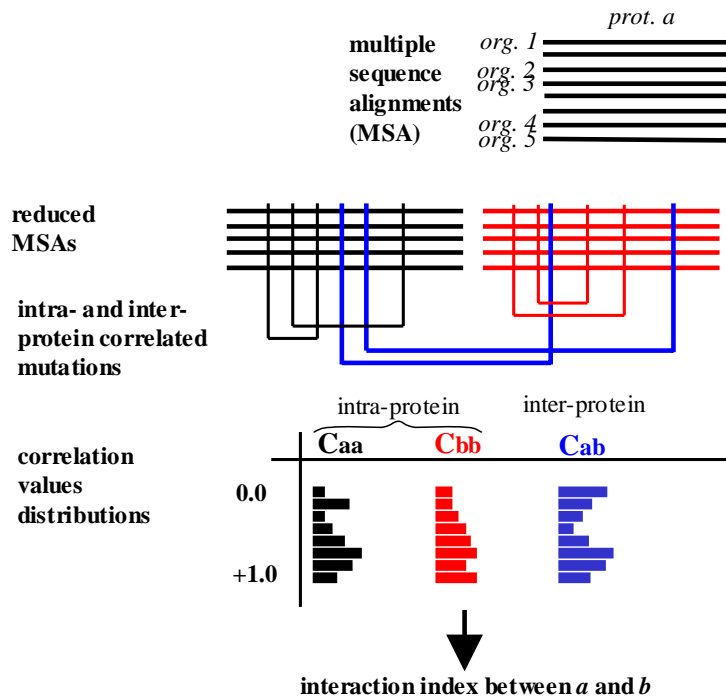
## b) conservation of gene neighbouring



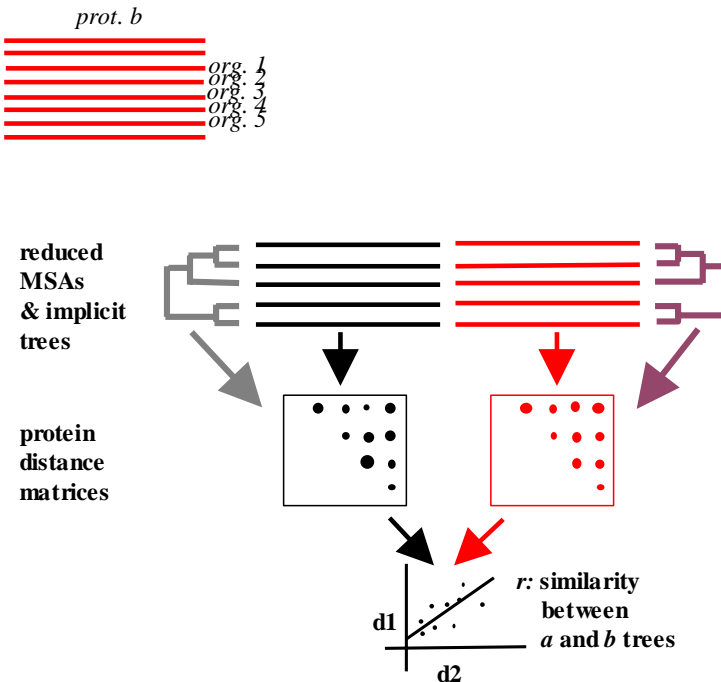
## c) gene fusion



## e) correlated mutations



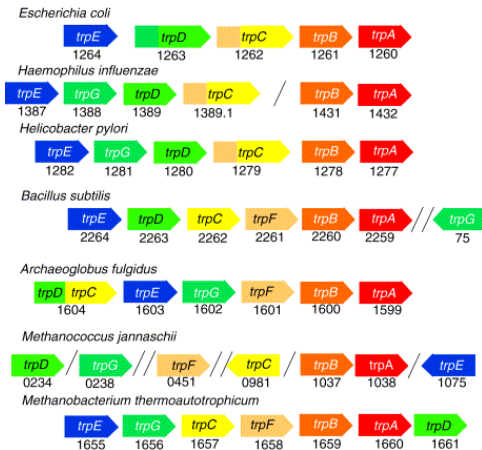
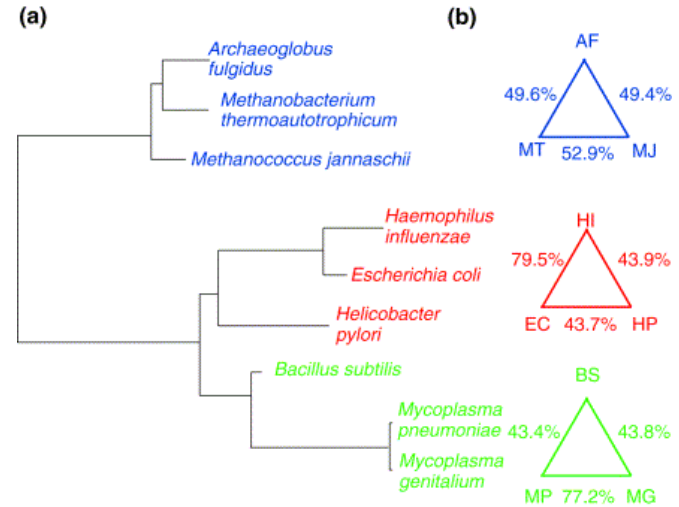
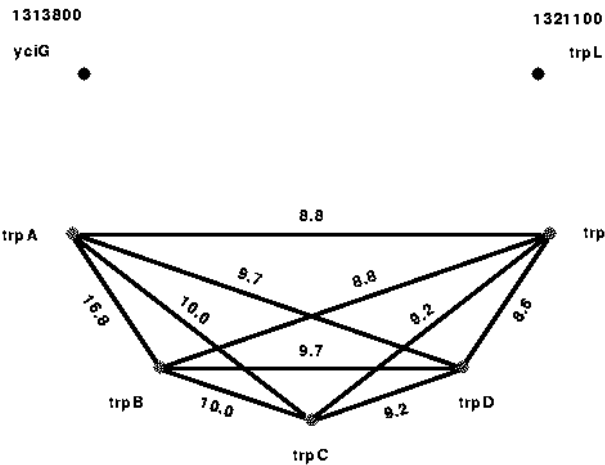
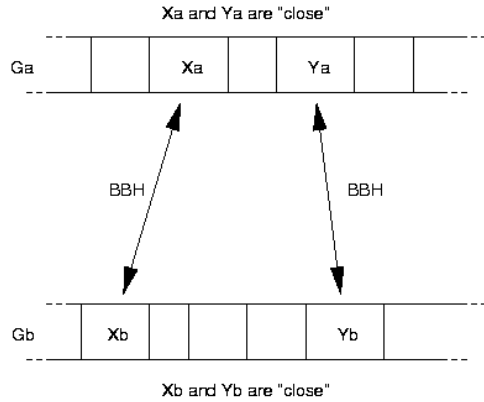
## d) similarity of phylogenetic trees



Huynen, M., Snel, B., Lathe, W. & Bork, P. (2000) Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res*, **10**, 1204-1210.

Valencia, A. & Pazos, F. (2002) Computational methods for the prediction of protein interactions. *Curr Opin Struct Biol*, **12**, 368-373.

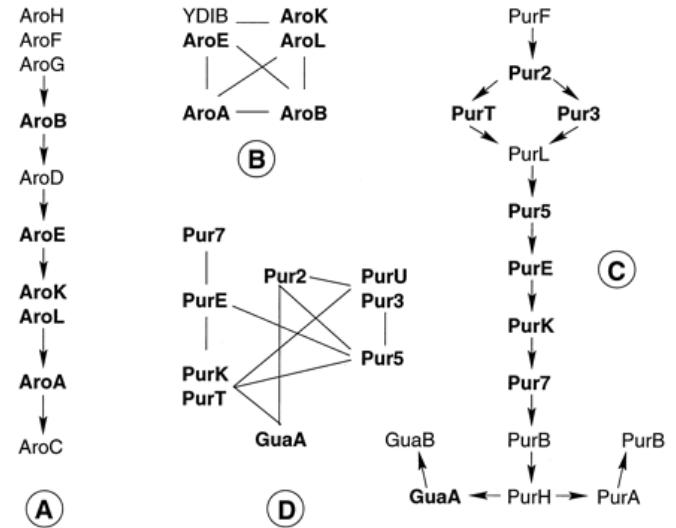
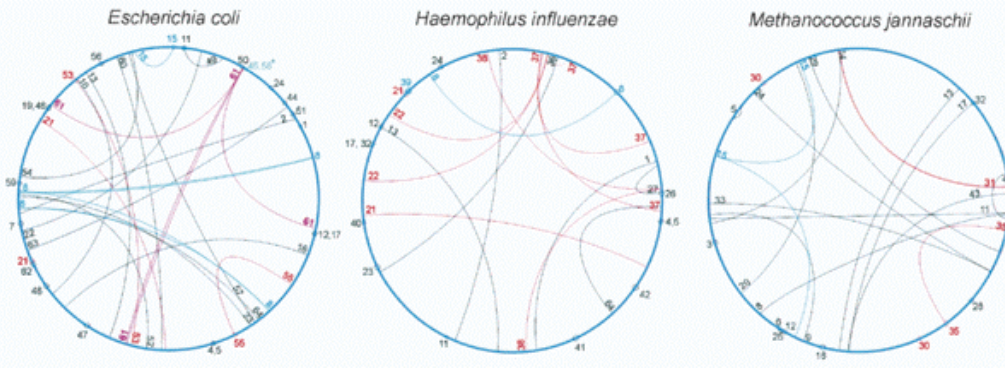
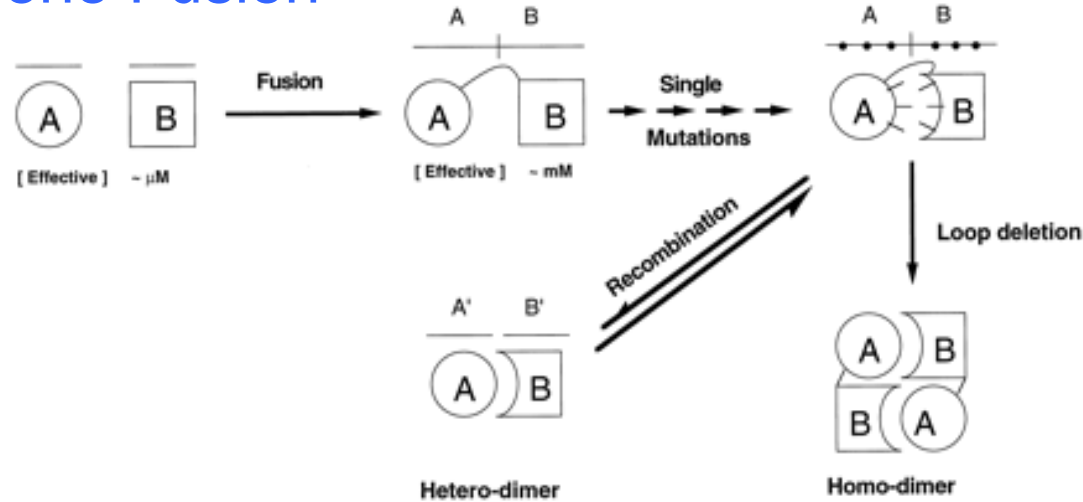
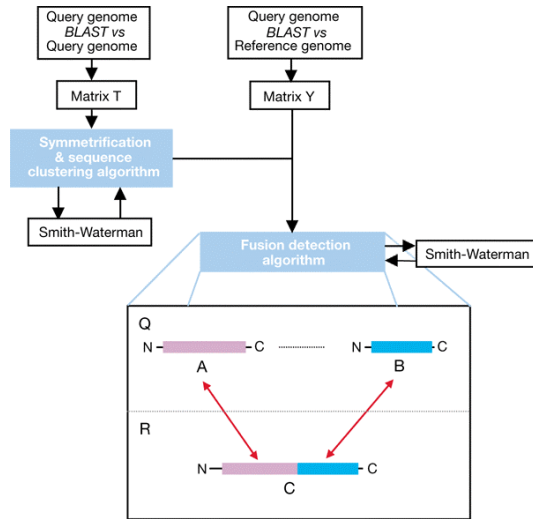
# Conservation of gene neighboring



Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D. & Maltsev, N. (1999). Use of contiguity on the chromosome to predict functional coupling. *In Silico Biol.* **1**, 93-108.

Dandekar, T., Snel, B., Huynen, M. & Bork, P. (1998). Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci.* **23**, 324-328.

# Gene Fusion

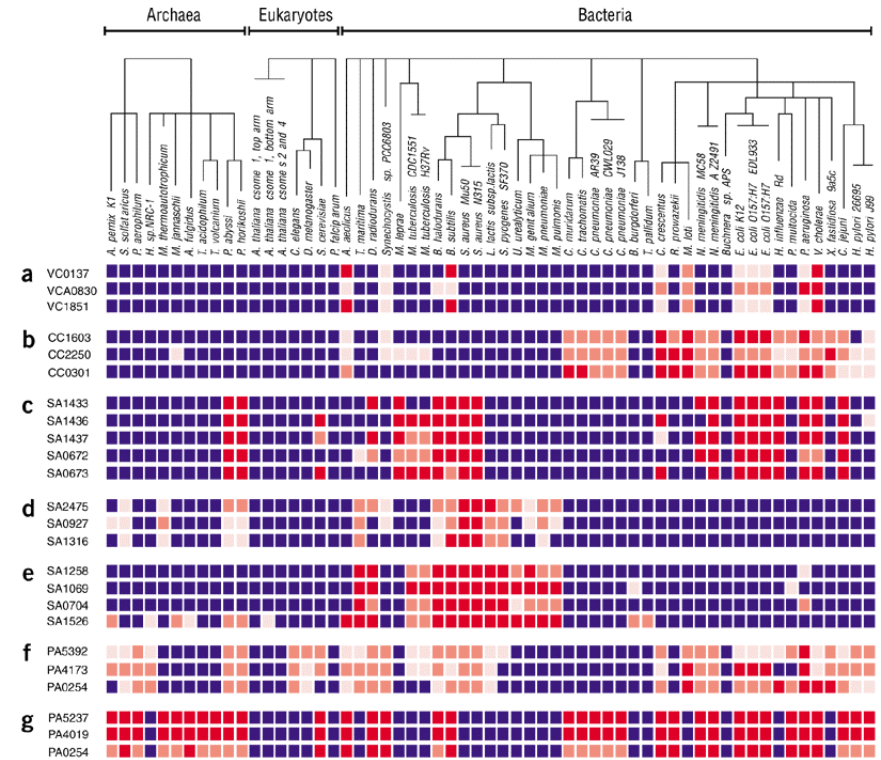
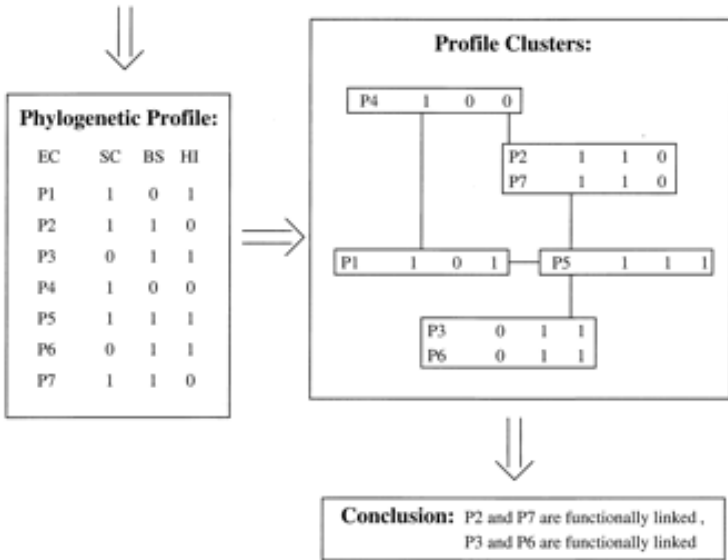
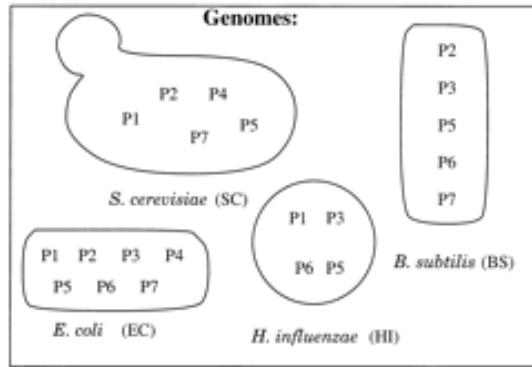


Enright, A. J., Iliopoulos, I., Kyripides, N. C. & Ouzounis, C. A. (1999). Protein interaction maps for complete genomes based on gene fusion events. *Nature*. **402**, 86-90.

Marcotte, E. M., Pellegrini, M., Ho-Leung, N., Rice, D. W., Yeates, T. O. & Eisenberg, D. (1999). Detecting protein function and protein-protein interactions from genome sequences. *Science*. **285**, 751-753.



# Phylogenetic Profiles



$$p_{ij} = -1/\log E_{ij}$$

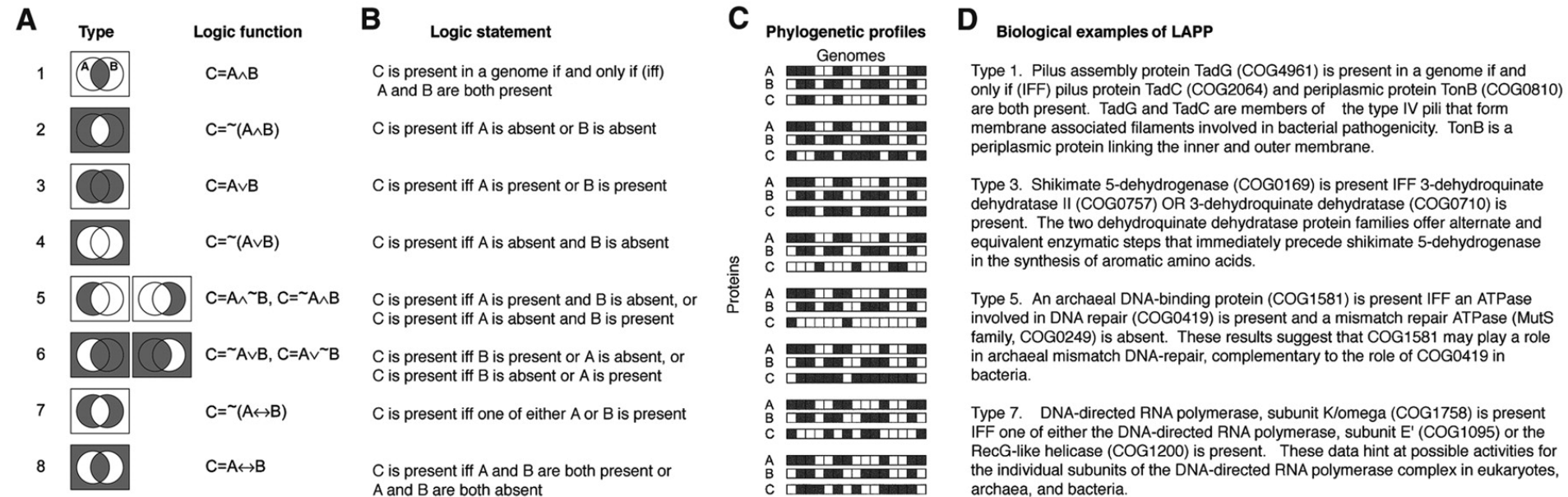
$$H(A) = - \sum p(a) \ln p(a)$$

$$MI(A,B) = H(A) + H(B) - H(A,B)$$

•Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. (1999). Assigning protein functions by comparative genome analysis: Protein pylogenetic profiles. *Proc Natl Acad Sci USA*. **96**, 4285-4288.

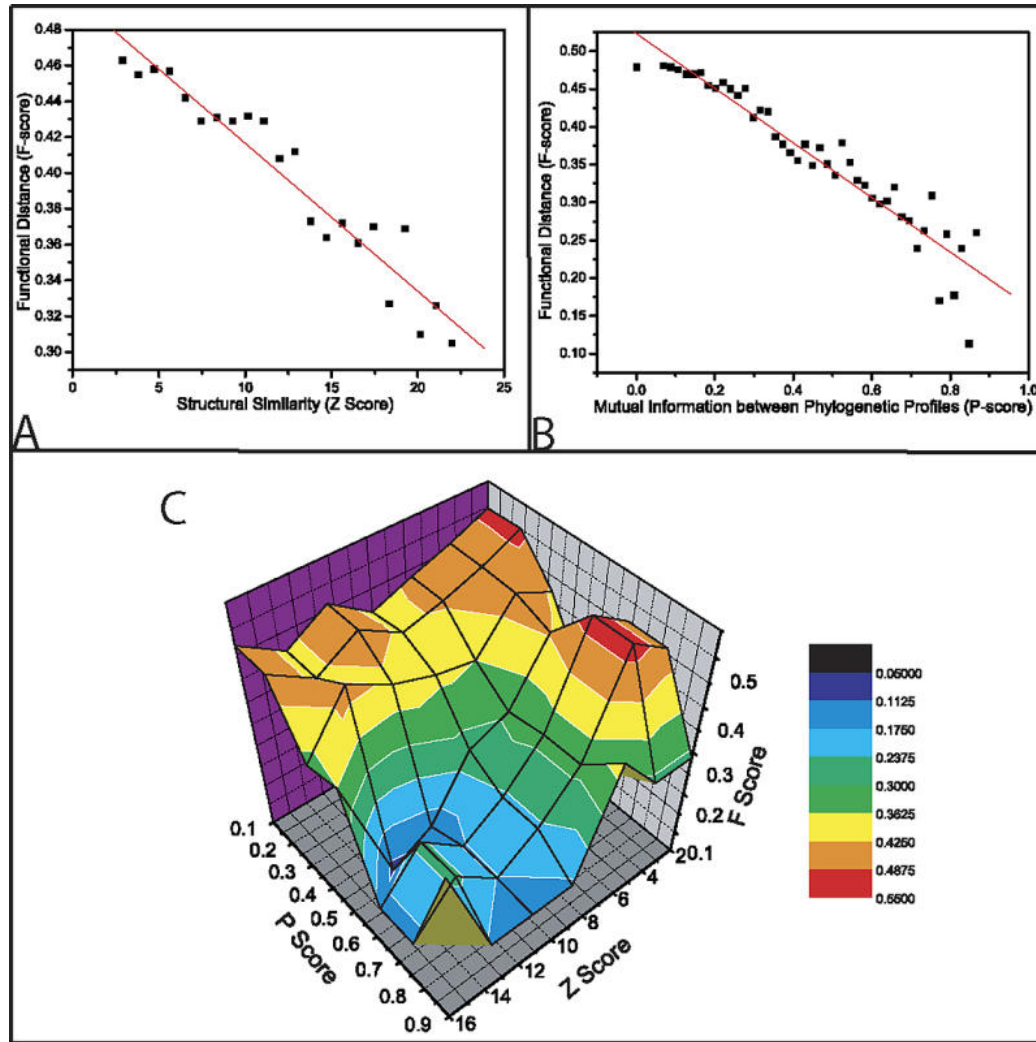
•Date, S. V. & Marcotte, E. M. (2003). Discovery of uncharacterized cellular systems by genome-wide analysis of functional linkages. *Nat Biotechnol*. **21**, 1055-1062.

# Phylogenetic Profiles



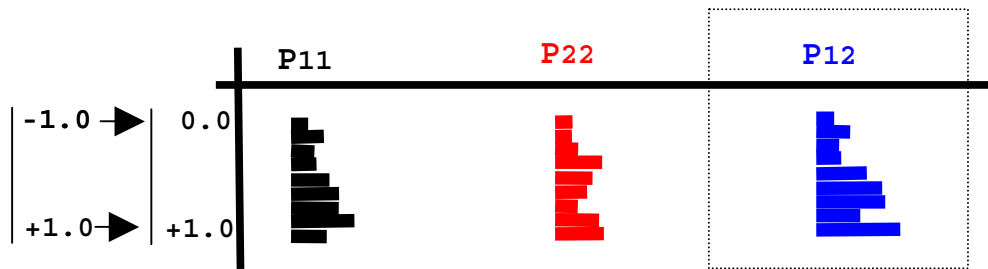
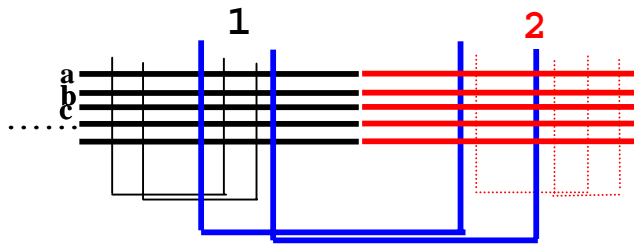
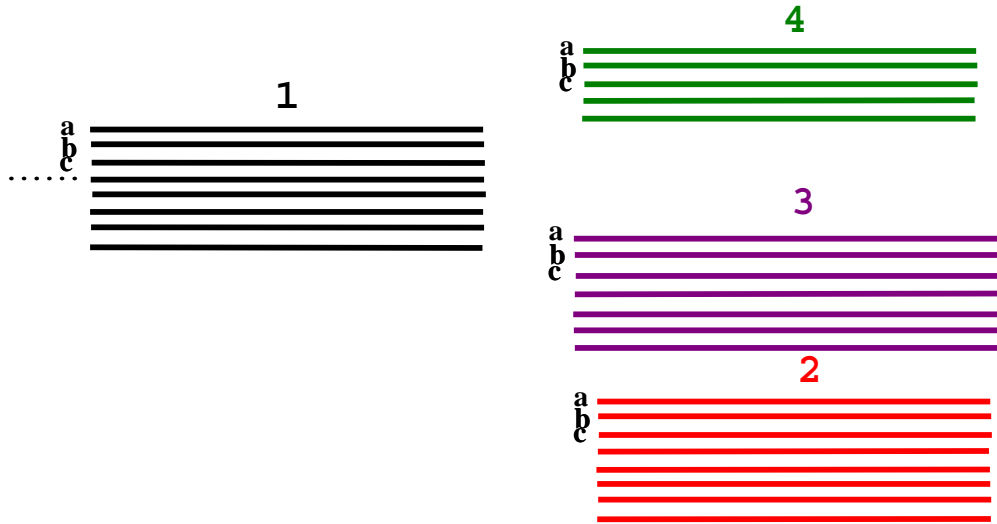


# Phylogenetic Profiles



Shakhnovich, B.E. (2005) Improving the precision of the structure-function relationship by considering phylogenetic context. *PLoS Comput Biol*, 1, e9.

# Based on correlated mutations (*i2h*)



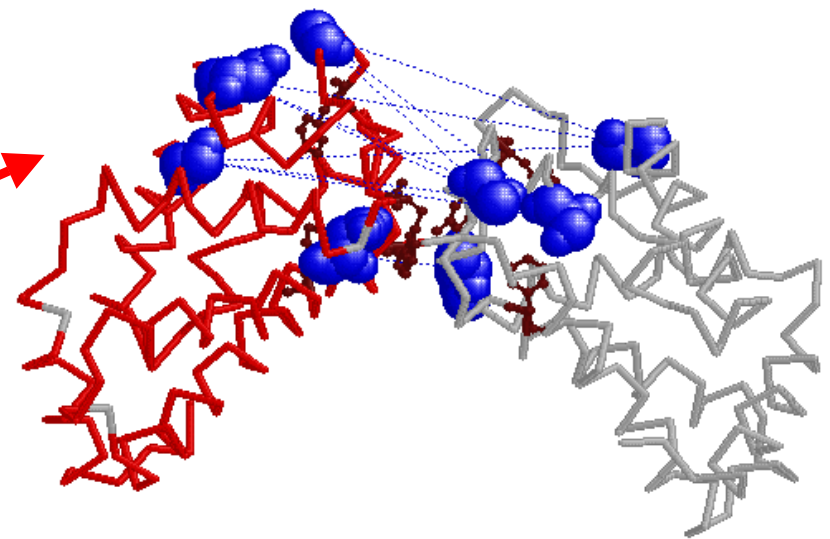
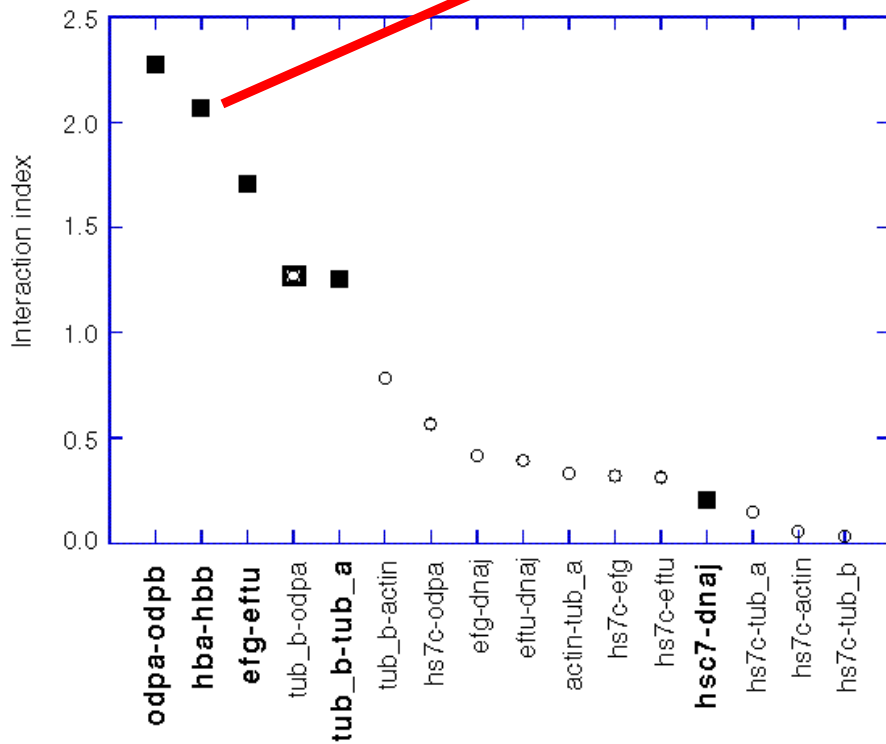
$$C_{12} = \sum_{i=incorr}^{1.0} \frac{P_{12i}}{P_{11i} + P_{22i}} \cdot i$$

# Based on correlated mutations (i2h)

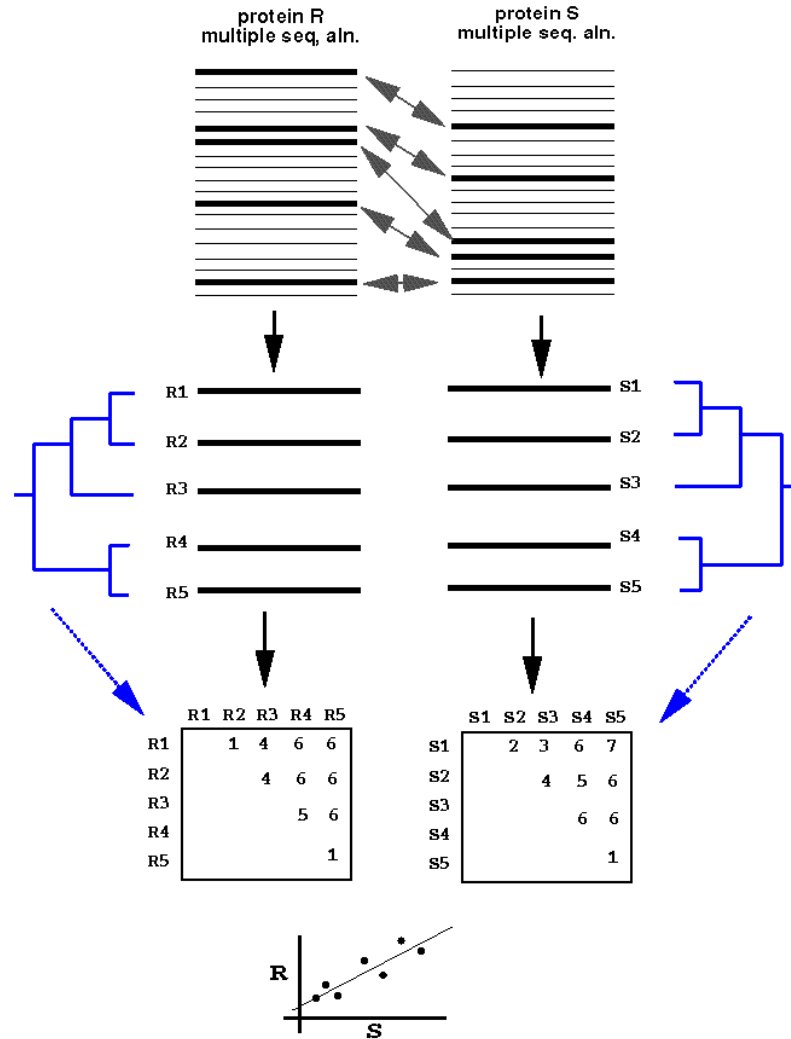
2c2c_2-1alc_1		3,503	1alc_1-1rnd_1		0,968
1sgt_2-4mt2_1		3,448	3adk_2-4tnc_2		0,961
<b>9pap_1-9pap_2</b>	*	<b>3,042</b>	1alc_1-1rnd_2		0,957
<b>1alc_1-1alc_2</b>	*	<b>2,852</b>	1sgt_1-2c2c_2		0,889
2c2c_1-4mt2_1		2,825	2c2c_2-3pgk_1		0,878
<b>4tms_1-4tms_2</b>	*	<b>2,735</b>	3trx_1-9pap_2		0,857
<b>3trx_1-3trx_2</b>	*	<b>2,571</b>	4tnc_1-4mt2_2		0,853
<b>4mt2_1-4mt2_2</b>	*	<b>2,469</b>	4tnc_2-4mt2_2		0,836
2c2c_2-4mt2_1		2,355	3trx_1-3pgk_2		0,829
2c2c_2-4mt2_2		2,331	3trx_1-9pap_1		0,814
<b>4tnc_1-4tnc_2</b>	*	<b>2,238</b>	2c2c_2-1rnd_2		0,813
<b>3blm_1-3blm_2</b>	*	<b>2,206</b>	4tms_2-3dfr_2		0,809
<b>3pgk_1-3pgk_2</b>	*	<b>2,197</b>	9pap_2-3adk_2		0,805
2c2c_1-4mt2_2		2,139	4tms_1-3dfr_2		0,804
1sgt_2-2c2c_1		2,068	1sgt_2-1alc_1		0,799
2c2c_1-1alc_1		2,011	9pap_1-3adk_2		0,790
2c2c_1-1alc_2		1,886	3trx_2-9pap_2		0,761
<b>3adk_1-3adk_2</b>	*	<b>1,862</b>	4tnc_2-4mt2_1		0,747
1sgt_2-2c2c_2		1,835	3adk_2-3pgk_2		0,726
<b>2c2c_1-2c2c_2</b>	*	<b>1,787</b>	4tnc_1-4mt2_1		0,718
3adk_1-3pgk_1		1,624	9pap_2-4tnc_2		0,702
1rnd_1-4mt2_1		1,530	3trx_1-3adk_1		0,673
2c2c_1-9pap_2		1,520	<b>3dfr_1-3dfr_2</b>	*	<b>0,657</b>
3adk_2-3dfr_2		1,507	2pf2_2-1alc_2		0,628
1sgt_2-2pf2_2		1,489	3adk_1-4tnc_1		0,617
9pap_1-3adk_1		1,488	3adk_1-4tnc_2		0,614
3adk_1-3pgk_2		1,444	2pf2_2-1alc_1		0,595
2c2c_2-1alc_2		1,415	3adk_2-4tnc_1		0,539
2c2c_1-3pgk_2		1,389	4tms_1-3dfr_1		0,507
1sgt_1-4mt2_1		1,387	3trx_2-3pgk_1		0,489
3adk_1-3dfr_1		1,367	3trx_2-3pgk_2		0,471
1rnd_2-4mt2_1		1,359	3trx_1-3adk_2		0,471
2c2c_2-3adk_1		1,319	1sgt_1-1alc_1		0,455
<b>1rnd_1-1rnd_2</b>	*	<b>1,314</b>	3trx_1-2c2c_2		0,453
3pgk_1-4tms_1		1,299	3trx_1-2c2c_1		0,446
2c2c_1-3adk_1		1,297	4tms_2-4tnc_2		0,444
3pgk_1-4tms_2		1,292	2c2c_1-1rnd_2		0,442
3trx_1-3pgk_1		1,279	1sgt_2-1alc_2		0,435
2c2c_1-3pgk_1		1,278	3trx_2-3adk_1		0,427
1alc_1-4mt2_1		1,278	4tms_1-4tnc_2		0,413
2c2c_2-9pap_2		1,274	1sgt_1-1rnd_1		0,403
1rnd_1-4mt2_2		1,258	4tms_1-4tnc_1		0,401
3adk_2-3pgk_1		1,252	4tms_2-3dfr_1		0,398
1rnd_2-4mt2_2		1,240	1alc_2-4mt2_2		0,362
3adk_1-3dfr_2		1,209	1sgt_1-1rnd_2		0,358
3trx_2-2c2c_1		1,196	1sgt_1-4mt2_2		0,356
3pgk_2-4tms_2		1,178	1sgt_2-1rnd_1		0,352
3pgk_2-4tms_1		1,170	3trx_1-4tnc_2		0,316
			2c2c_1-4tnc_1		0,303



# Based on correlated mutations (*i2h*)



# MirrorTree

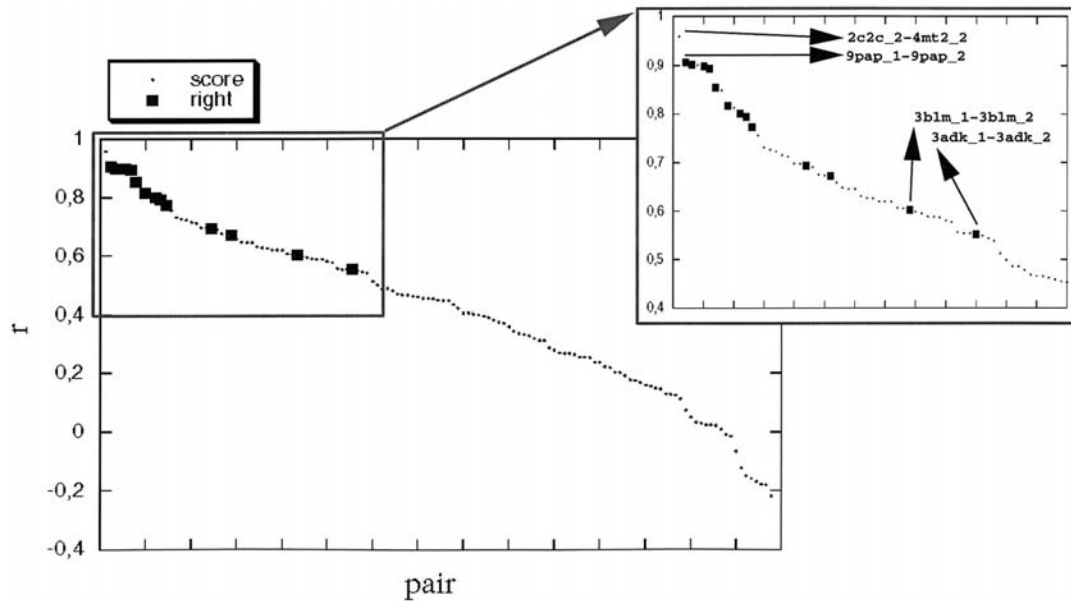


$$r = \frac{\sum_{i=1}^n (R_i - \bar{R}) \cdot (S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2} \cdot \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}}$$

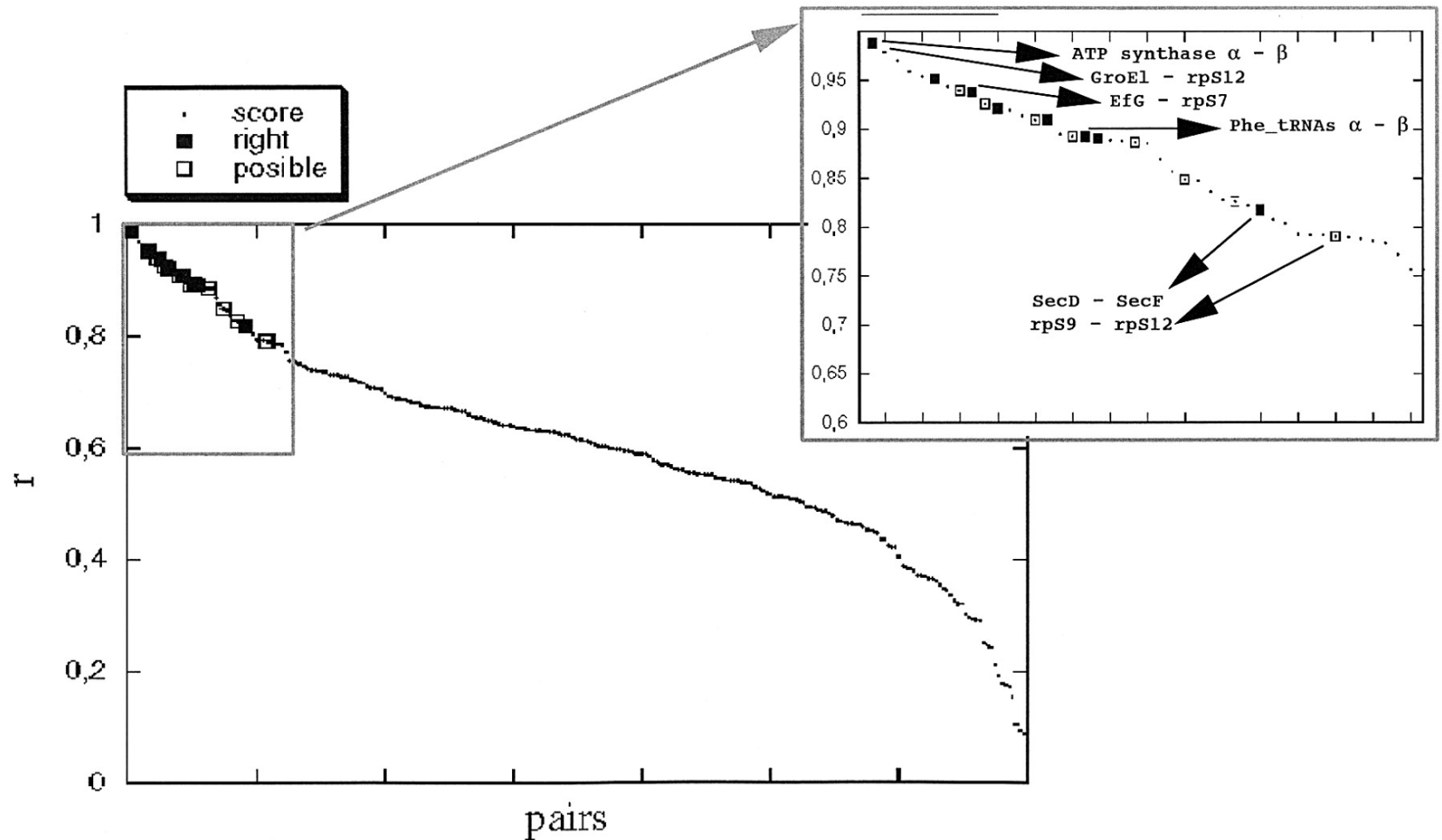
Goh, C.-S., Bogan, A.A., Joachimiak, M., Walther, D. and Cohen, F.E. (2000) Co-evolution of Proteins with their Interaction Partners. *J Mol Biol*, **299**, 283-293.

Pazos, F. and Valencia, A. (2001) Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Eng*, **14**, 609-614.

2c2c_2-4mt2_2	0,959	4tnc_1-4mt2_2	0,448
<b>9pap_1-9pap_2</b> *	<b>0,907</b>	9pap_2-4tnc_1	0,446
2c2c_1-4mt2_2	0,901	1alc_2-4mt2_2	0,446
<b>3pgk_1-3pgk_2</b> *	<b>0,901</b>	1sgt_2-4mt2_1	0,433
<b>4mt2_1-4mt2_2</b> *	<b>0,898</b>	3adk_1-4tnc_2	0,421
<b>3trx_1-3trx_2</b> *	<b>0,894</b>	1rnd_1-4mt2_2	0,405
<b>4tms_1-4tms_2</b> *	<b>0,854</b>	4tnc_2-4mt2_2	0,405
2c2c_2-4mt2_1	0,849	2c2c_1-3adk_2	0,401
<b>1rnd_1-1rnd_2</b> *	<b>0,817</b>	1sgt_2-2c2c_1	0,399
2c2c_1-4mt2_1	0,813	4tms_2-3dfr_2	0,394
<b>1alc_1-1alc_2</b> *	<b>0,801</b>	3adk_1-3dfr_1	0,390
<b>4tnc_1-4tnc_2</b> *	<b>0,794</b>	1sgt_2-2c2c_2	0,381
<b>2c2c_1-2c2c_2</b> *	<b>0,773</b>	3adk_2-3dfr_1	0,372
3pgk_1-4tms_1	0,756	1sgt_2-1alc_1	0,371
3pgk_1-4tms_2	0,731	4tms_1-3dfr_2	0,358
2c2c_1-3adk_1	0,726	1sgt_1-4mt2_1	0,343
3pgk_2-4tms_1	0,723	1sgt_1-4mt2_2	0,336
2c2c_2-3pgk_1	0,715	9pap_2-4tnc_2	0,331
1alc_1-1rnd_1	0,712	4tms_1-3dfr_1	0,327
2c2c_2-3pgk_2	0,698	3trx_1-2c2c_2	0,319
1alc_2-1rnd_1	0,697	3trx_1-2c2c_1	0,312
<b>1sgt_1-1sgt_2</b> *	<b>0,693</b>	1sgt_1-1alc_1	0,312
3pgk_2-4tms_2	0,691	3trx_2-2c2c_2	0,287
3adk_2-3dfr_2	0,675	3trx_2-2c2c_1	0,281
1sgt_2-2pf2_2	0,673	1sgt_1-2c2c_2	0,270
<b>3dfr_1-3dfr_2</b> *	<b>0,672</b>	1alc_1-4mt2_2	0,268
2c2c_2-9pap_1	0,658	1sgt_1-2c2c_1	0,268
2c2c_1-3pgk_1	0,648	2c2c_1-1rnd_1	0,263
3trx_2-9pap_1	0,646	9pap_1-3adk_2	0,254
1sgt_1-2pf2_2	0,646	2c2c_2-3adk_2	0,254
2c2c_2-3adk_1	0,631	3adk_2-3pgk_1	0,251
3trx_1-9pap_1	0,627	1sgt_1-1rnd_1	0,238
2c2c_2-1alc_2	0,626	3adk_2-3pgk_2	0,238
2c2c_1-3pgk_2	0,620	9pap_2-3adk_2	0,221
3trx_2-9pap_2	0,620	1sgt_2-1alc_2	0,219
1rnd_2-4mt2_1	0,619	2c2c_2-1alc_1	0,203
1alc_2-1rnd_2	0,607	9pap_1-4tnc_1	0,202
1rnd_2-4mt2_2	0,606	1sgt_2-1rnd_1	0,191
<b>3blm_1-3blm_2</b> *	<b>0,603</b>	1sgt_1-1alc_2	0,178
1alc_1-1rnd_2	0,599	3trx_2-3adk_2	0,175
3trx_1-3pgk_1	0,595	1sgt_1-1rnd_2	0,168
3trx_1-9pap_2	0,589	2pf2_2-1alc_1	0,160
1alc_2-4mt2_1	0,588	2c2c_1-1alc_1	0,155
2c2c_1-1alc_2	0,587	9pap_1-4tnc_2	0,149
2c2c_1-9pap_1	0,581	2c2c_2-1rnd_2	0,146
3trx_1-3pgk_2	0,577	4tms_2-3dfr_1	0,130
4tnc_1-4mt2_1	0,556	3trx_1-3adk_2	0,128
3adk_1-3pgk_1	0,554	2c2c_2-1rnd_1	0,125
		2c2c_1-1rnd_2	0,113







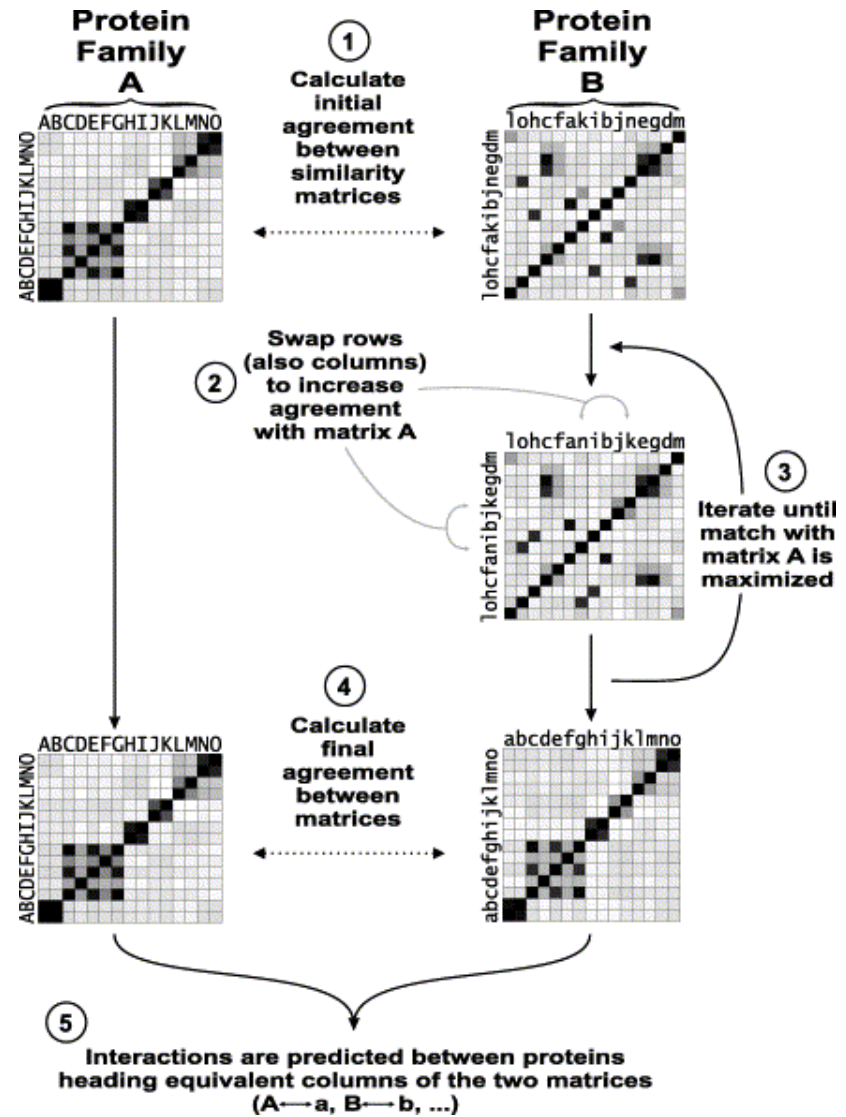
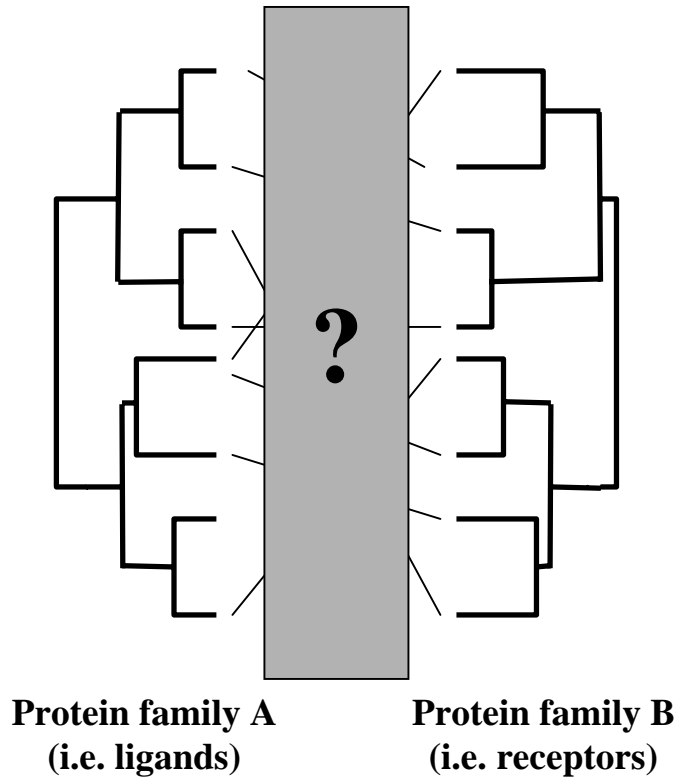
Coevolution of Permanent vs. Transient complexes:

Mintseris, J. and Weng, Z. (2005) Structure, function, and evolution of transient and obligate protein-protein interactions. *Proc Natl Acad Sci U S A*, **102**, 10930-10935.

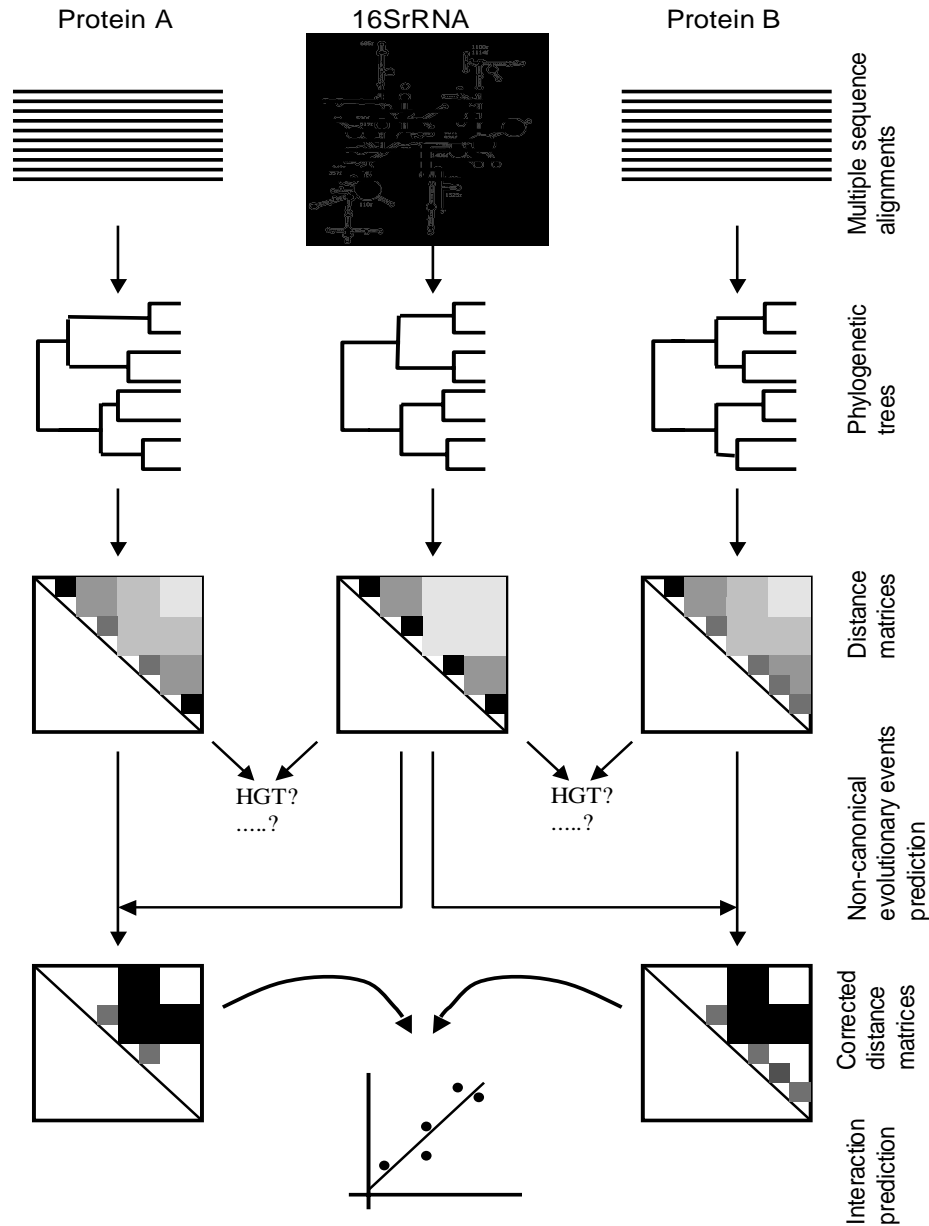
# MirrorTree Variations

- Gertz, J., Elfond, G., Shustrova, A., Weisinger, M., Pellegrini, M., Cokus, S. and Rothschild, B. (2003) Inferring protein interactions from phylogenetic distance matrices. *Bioinformatics*, **19**, 2039-2045.
- Goh, C.S. and Cohen, F.E. (2002) Co-evolutionary analysis reveals insights into protein-protein interactions. *J Mol Biol*, **324**, 177-192.
- Ramani, A.K. and Marcotte, E.M. (2003) Exploiting the co-evolution of interacting proteins to discover interaction specificity. *J Mol Biol*, **327**, 273-284.
- Sato, T., Yamanishi, Y., Horimoto, K., Toh, H. and Kanehisa, M. (2003) Prediction of protein-protein interactions from phylogenetic trees using partial correlation coefficient. *Genome Informatics*, **14**, 496-497.
- Kim, W.K., Bolser, D.M. and Park, J.H. (2004) Large-scale co-evolution analysis of protein structural interlogues using the global protein structural interactome map (PSIMAP). *Bioinformatics*, **20**, 1138-1150. Epub 2004 Feb 1135.
- Tan, S., Zhang, Z. and Ng, S. (2004) ADVICE: Automated Detection and Validation of Interaction by Co-Evolution. *Nucl. Acids. Res.*, **32**, W69-W72.
- Jothi, R., Kann, M.G. and Przytycka, T.M. (2005) Predicting protein-protein interaction by searching evolutionary tree automorphism space. *Bioinformatics*, **21**, i241-i250.
- Mintseris, J. and Weng, Z. (2005) Structure, function, and evolution of transient and obligate protein-protein interactions. *Proc Natl Acad Sci U S A*, **102**, 10930-10935.
- Sato, T., Yamanishi, Y., Kanehisa, M. and Toh, H. (2005) The inference of protein-protein interactions by co-evolutionary analysis is improved by excluding the information about the phylogenetic relationships. *Bioinformatics*, **21**, 3482-3489.

# MirrorTree. Variations



# tol-mirrortree



$$r_{AB} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (dA_{ij}' - \bar{dA}^i) \cdot (dB_{ij}' - \bar{dB}^i)}{\sqrt{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (dA_{ij}' - \bar{dA}^i)^2} \cdot \sqrt{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (dB_{ij}' - \bar{dB}^i)^2}}$$

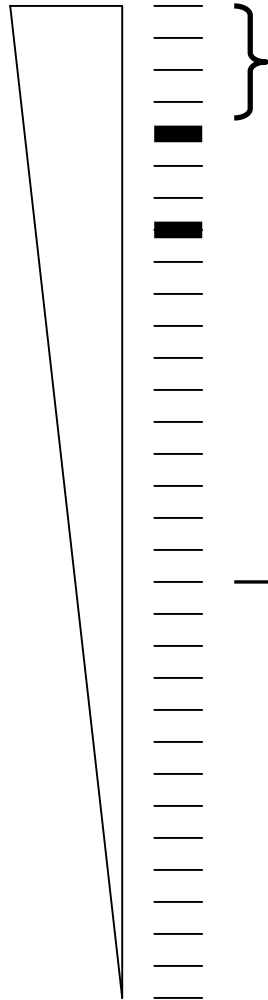
# Testing the method

DIP:  
516 interactions (*E coli*)

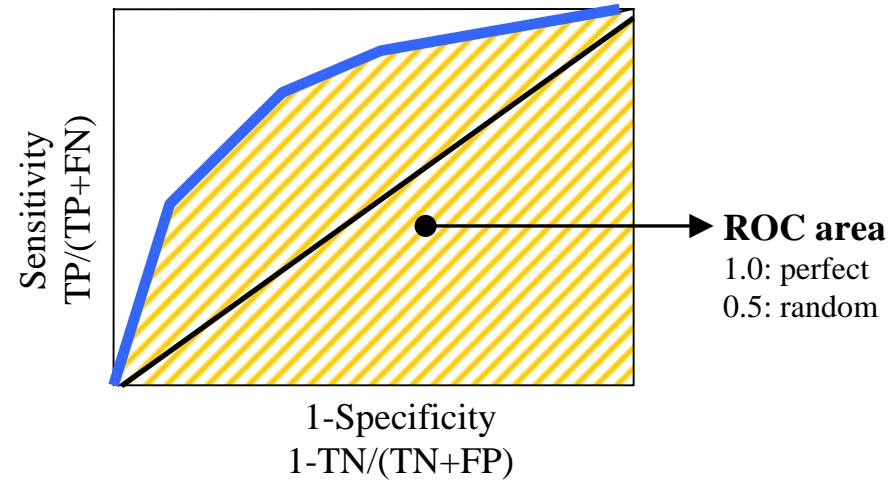
20,087 pairs calculated  
(115 true)

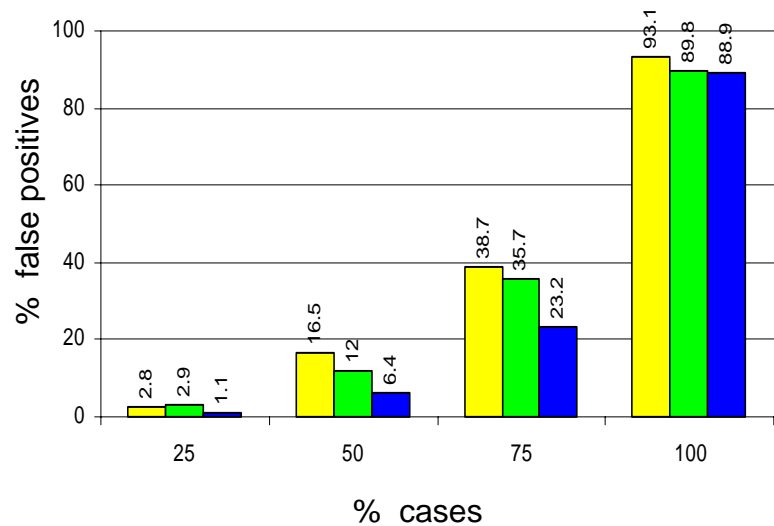
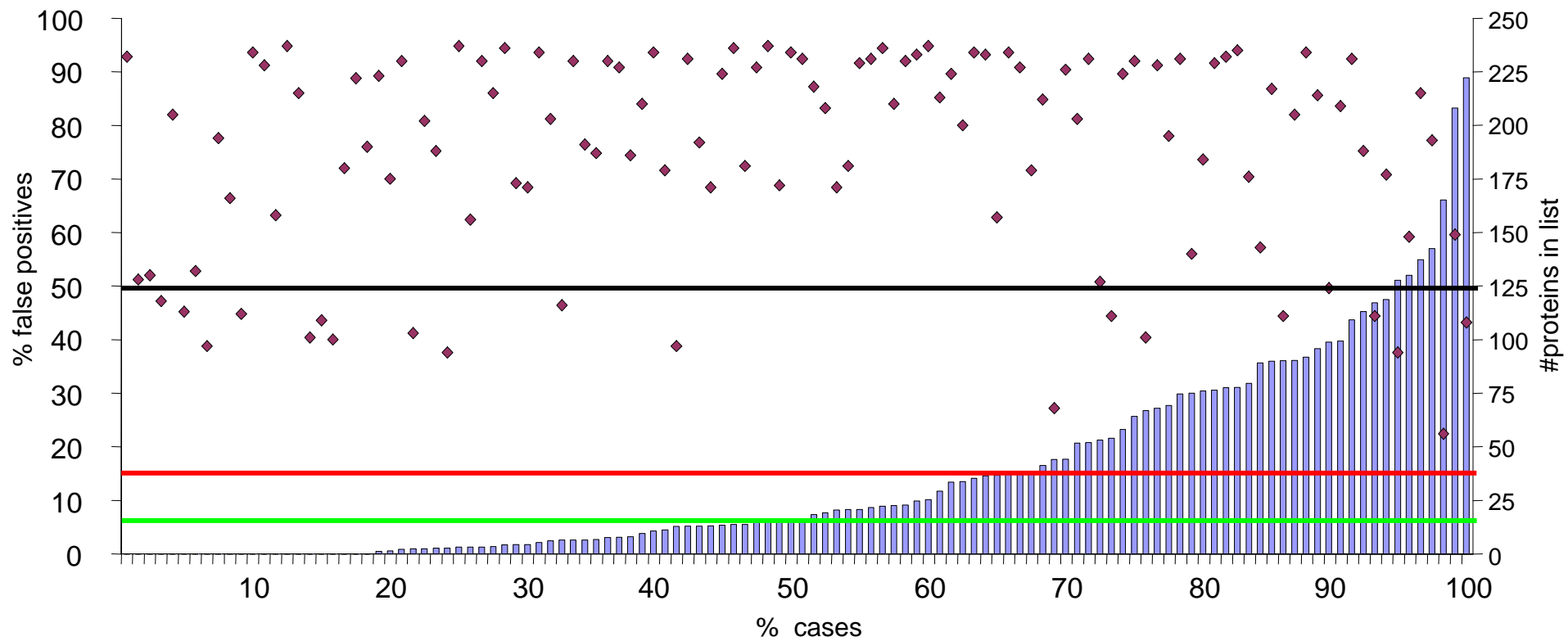
118 proteins with  
 $\geq 1$  calculated true interactor


P00000  
List of pairs  
sorted by score

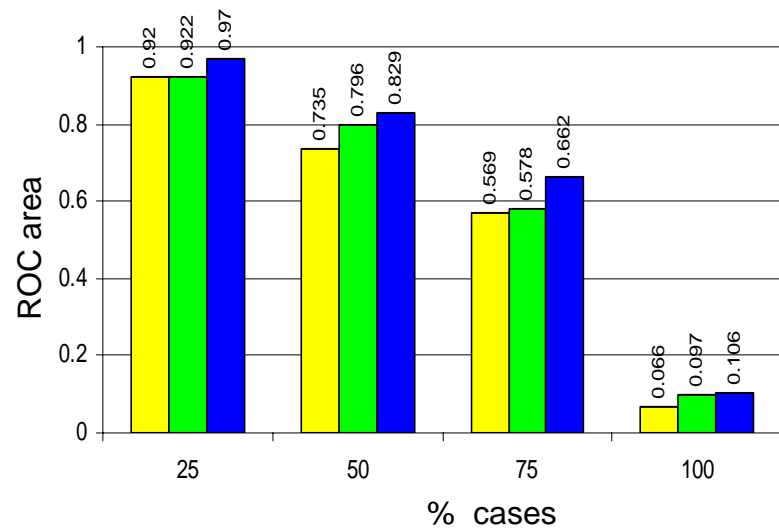


fraction of  
false positives  
0%: perfect  
50%: random  
(1 int.)





*mirrotree*  
  
*mirrotree*  
 (tree dist.)  
  
*tol-mirrotree*  

# Comparison with *old* methods

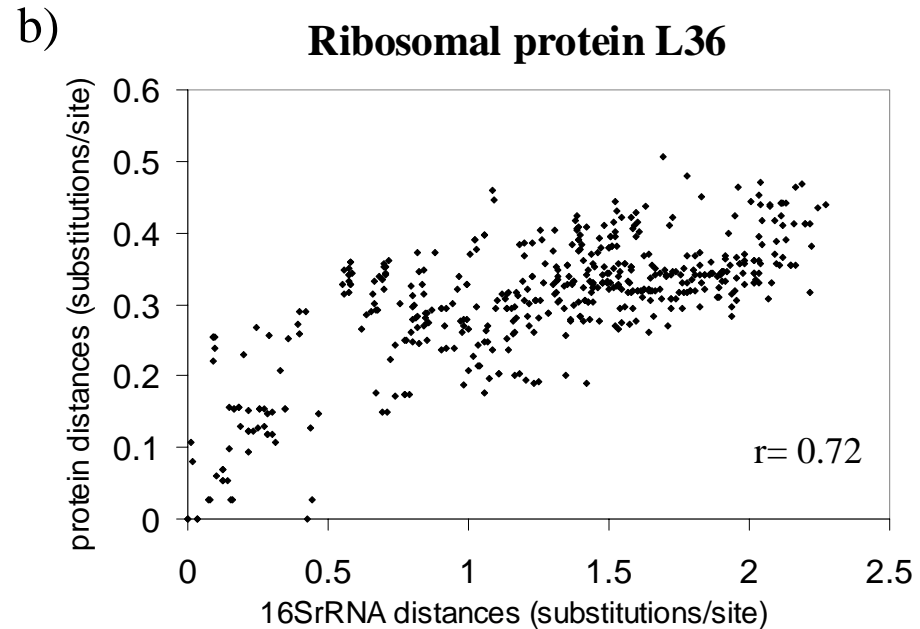
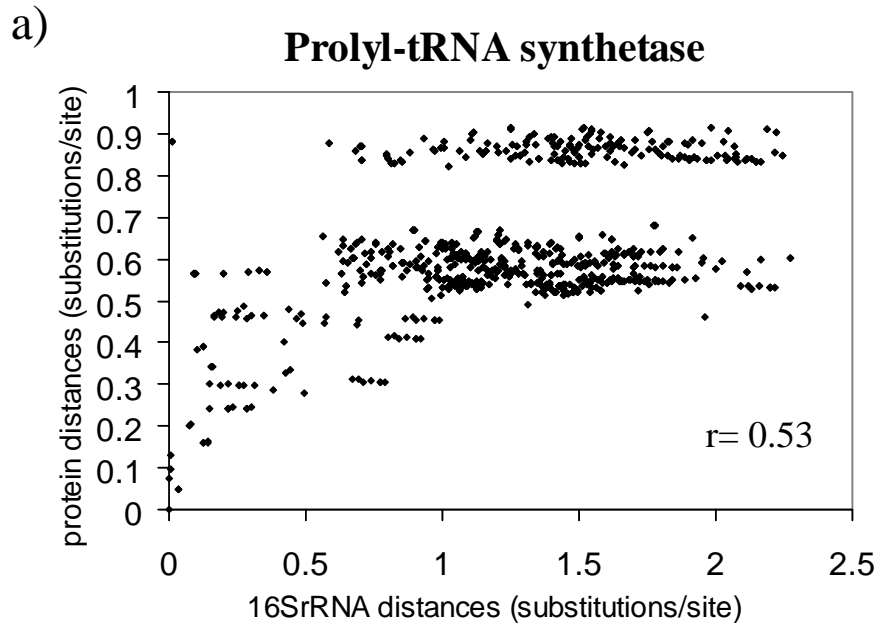
	% false positives	ROC area
<i>mirrortree</i>	23.4	0.71
<i>mirrortree</i> with tree distances	21.9	0.73
<b><i>tol-mirrortree</i></b>	<b>14.9</b>	<b>0.79</b>

**P(N) values (sign test):**

a \ b	<i>mirrortree</i>	<i>mirrortree</i> tree dist.	<i>tol-mirrortree</i>
<i>mirrortree</i>			
<i>mirrortree</i> tree dist.	0.276		
<i>tol-mirrortree</i>	$5.60 \cdot 10^{-6}$	$1.91 \cdot 10^{-5}$	

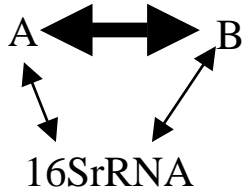


# Concomitant detection of non-standard evolutionary events

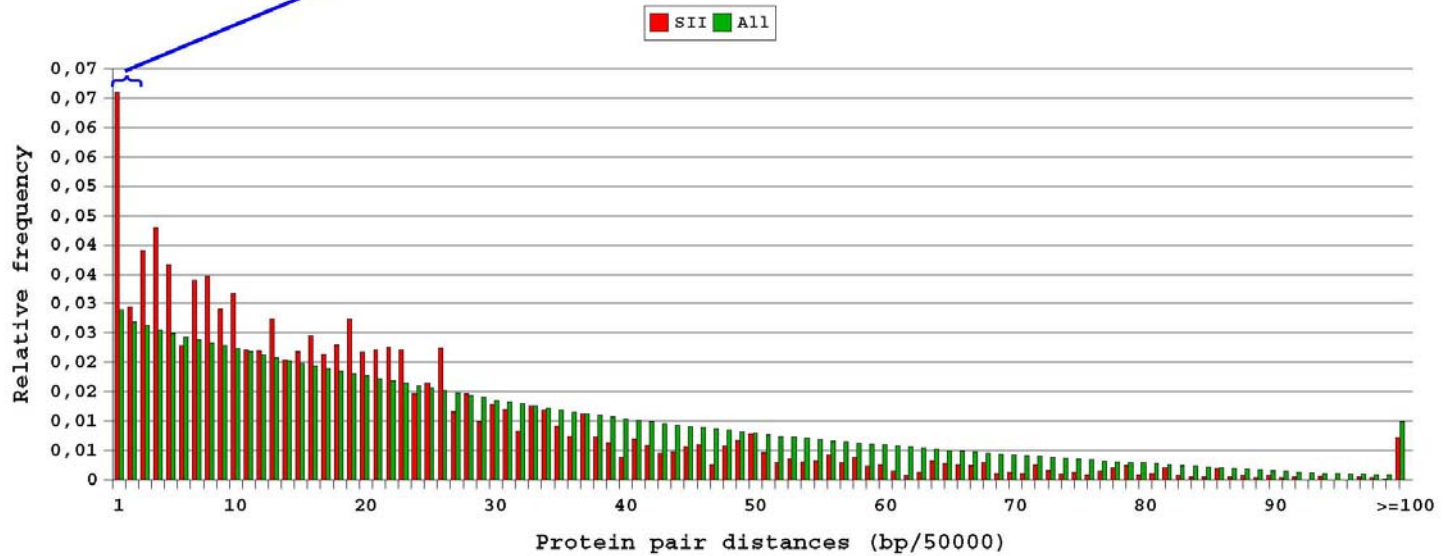
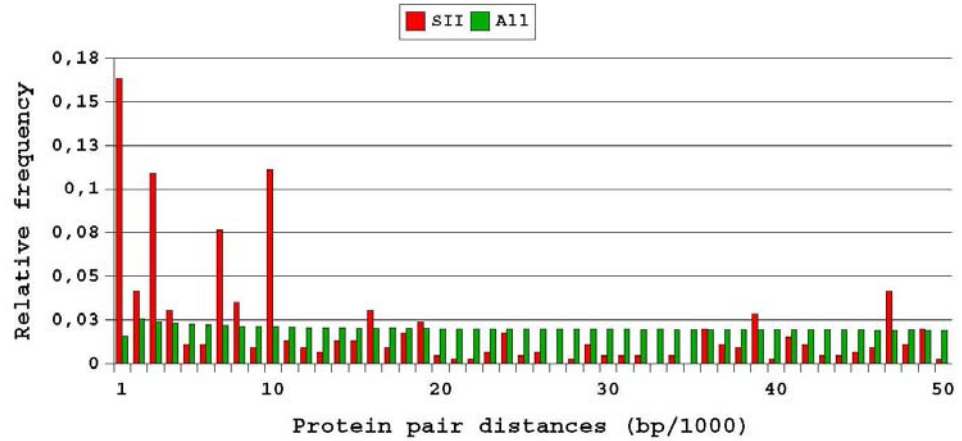


$r \leq 0.5 \rightarrow 25\%$  false pos (vs. 15%)  
Excluding them: 13.7% false pos (vs. 15%)

# Co-HGT events

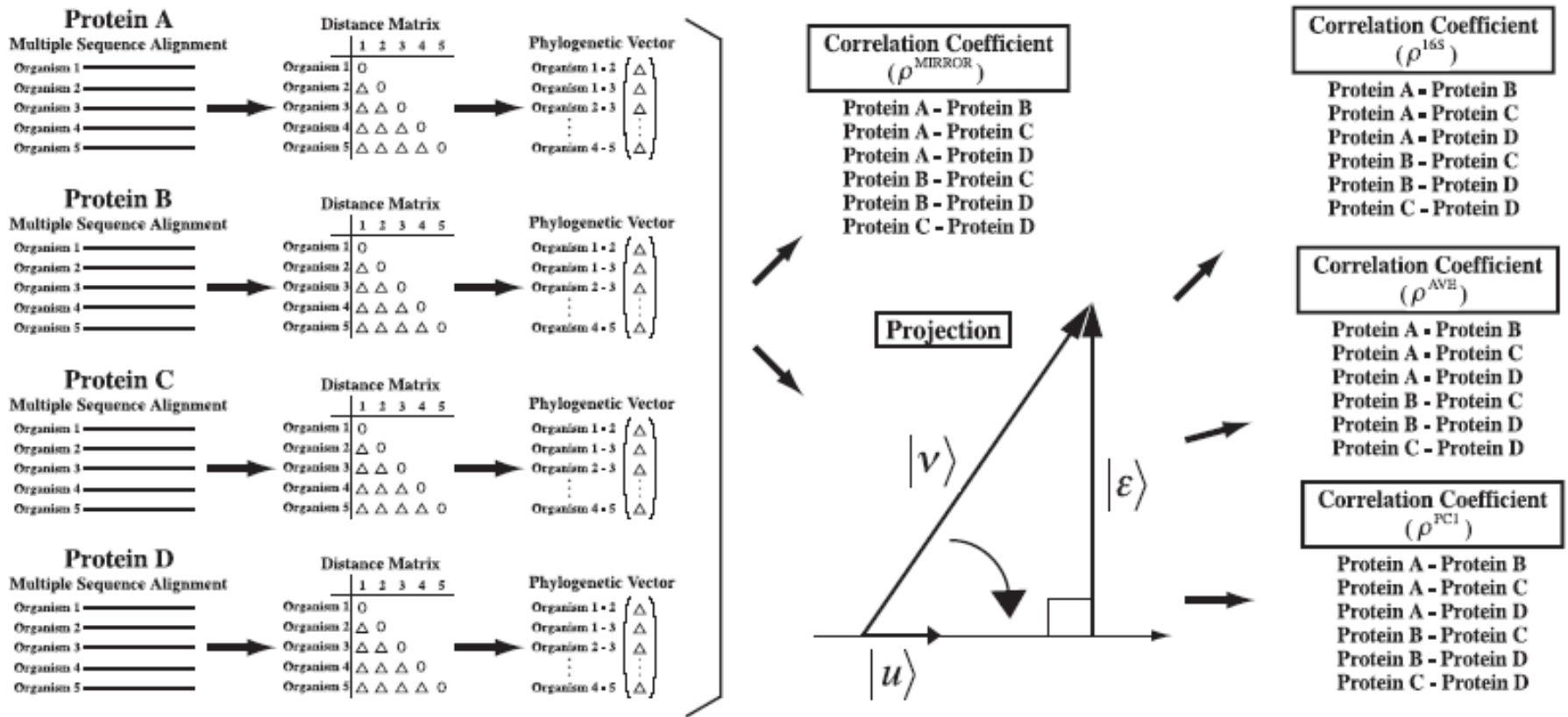


SII vs All *E. coli* protein orthologous pair distances



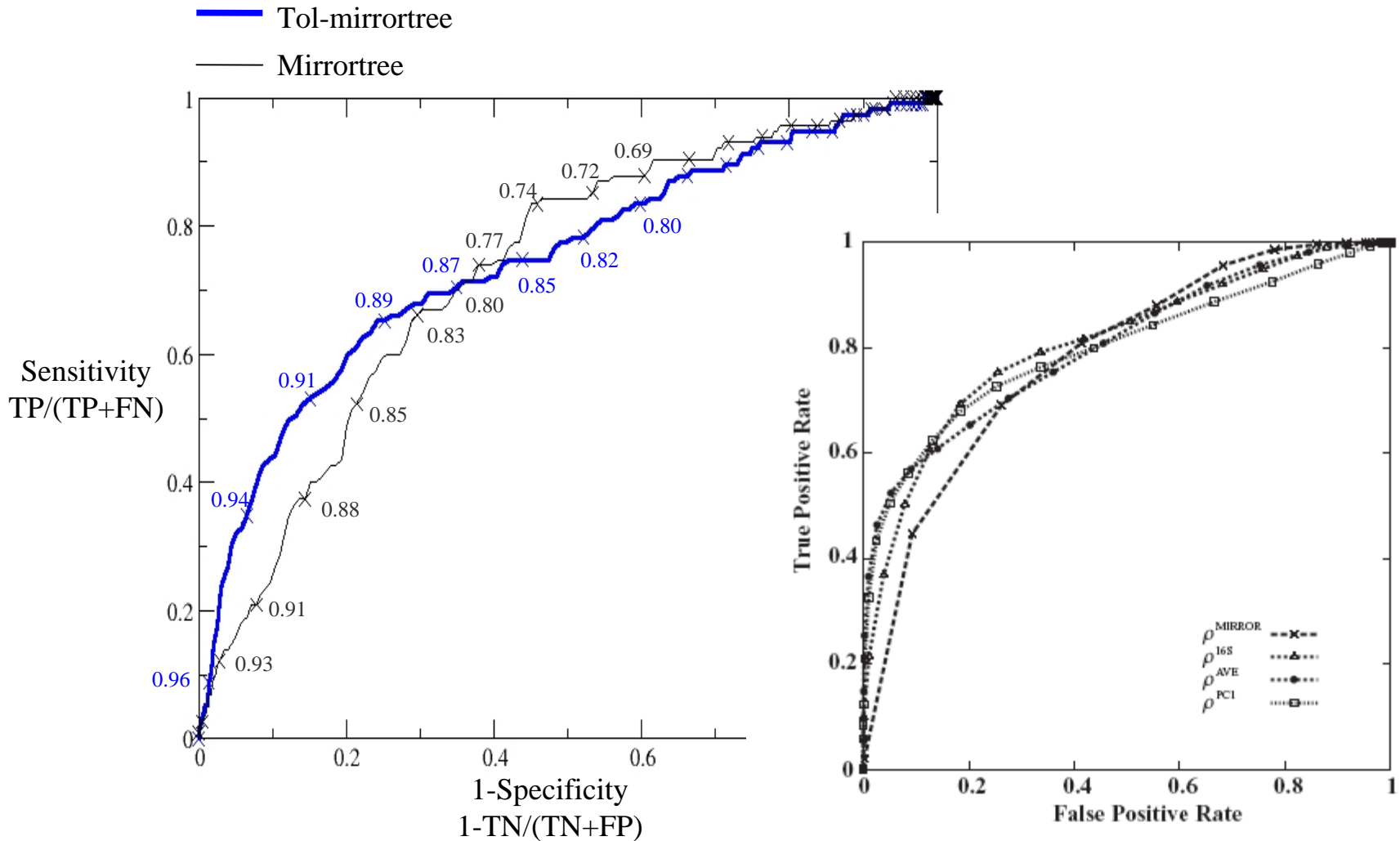
# tol-mirrortree

## Similar approaches



Sato, T., Yamanishi, Y., Kanehisa, M. and Toh, H. (2005) The inference of protein-protein interactions by co-evolutionary analysis is improved by excluding the information about the phylogenetic relationships. *Bioinformatics*, **21**, 3482-3489.

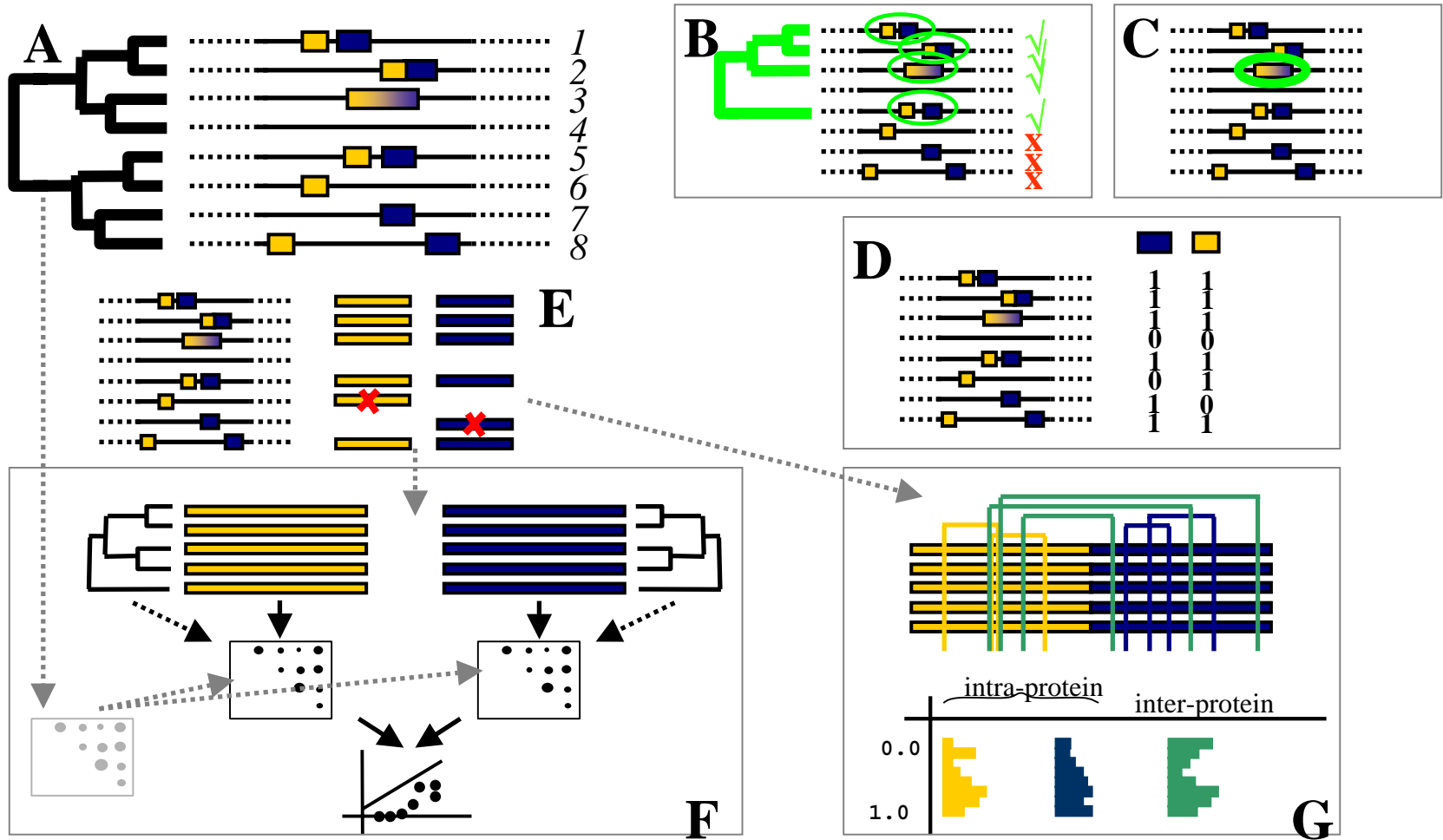
# tol-mirrortree



• Pazos, F., Ranea, J.A.G., Juan, D. and Sternberg, M.J.E. (2005) Assessing Protein Co-evolution in the Context of the Tree of Life Assists in the Prediction of the Interactome. *J Mol Biol*, **352**, 1002-1015.

• Sato, T., Yamanishi, Y., Kanehisa, M. and Toh, H. (2005) The inference of protein-protein interactions by co-evolutionary analysis is improved by excluding the information about the phylogenetic relationships. *Bioinformatics*, **21**, 3482-3489.

# Computational Methods for Predicting Interaction Partners - Overview

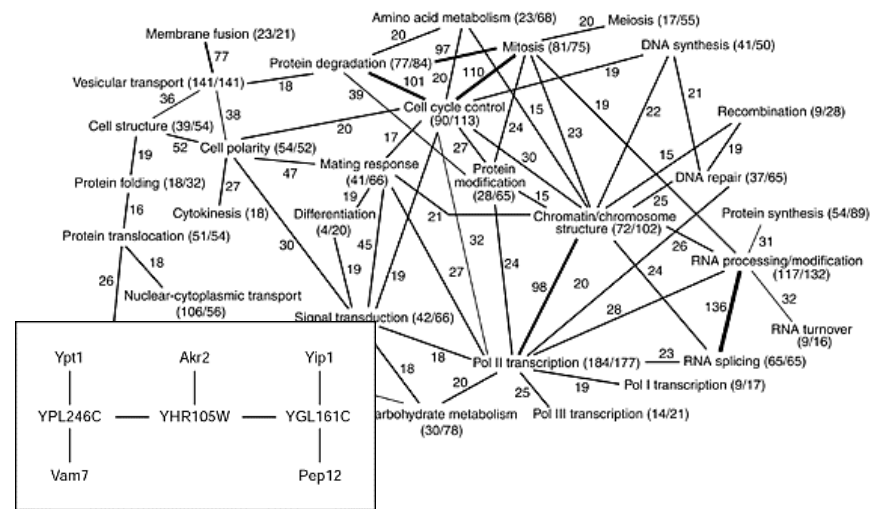
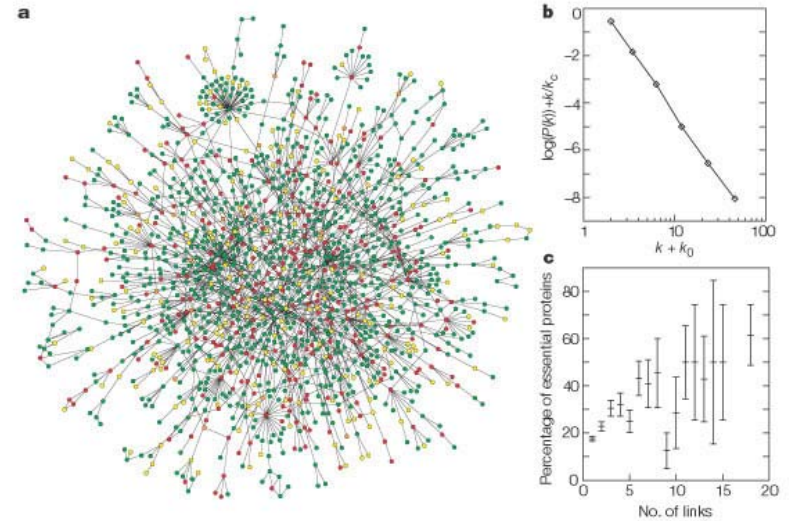
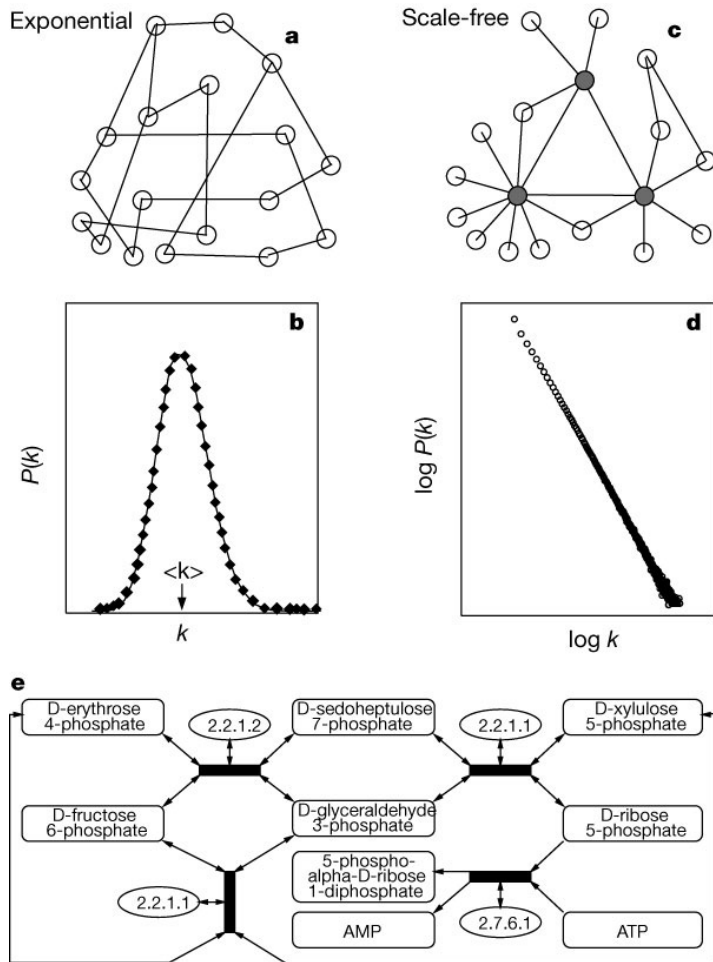


• Alfonso Valencia & Florencio Pazos (2002). Prediction of Protein Interactions with Computational Methods. *Curr Op Str Biol.* **12(3)**: 368-373. [56/67]

• Alfonso Valencia & Florencio Pazos. (2003). Prediction of protein-protein interactions from evolutionary information. *Methods Biochem Anal.* **44**:411-426.

• Florencio Pazos & Alfonso Valencia (2006). Protein Interactions from an Evolutionary Perspective. In "Evolution of Biological Networks". Carsten Wiuf & Michael Stumpf (Eds). Imperial College Press/World Scientific. *In Press*.

# System properties of the interactome

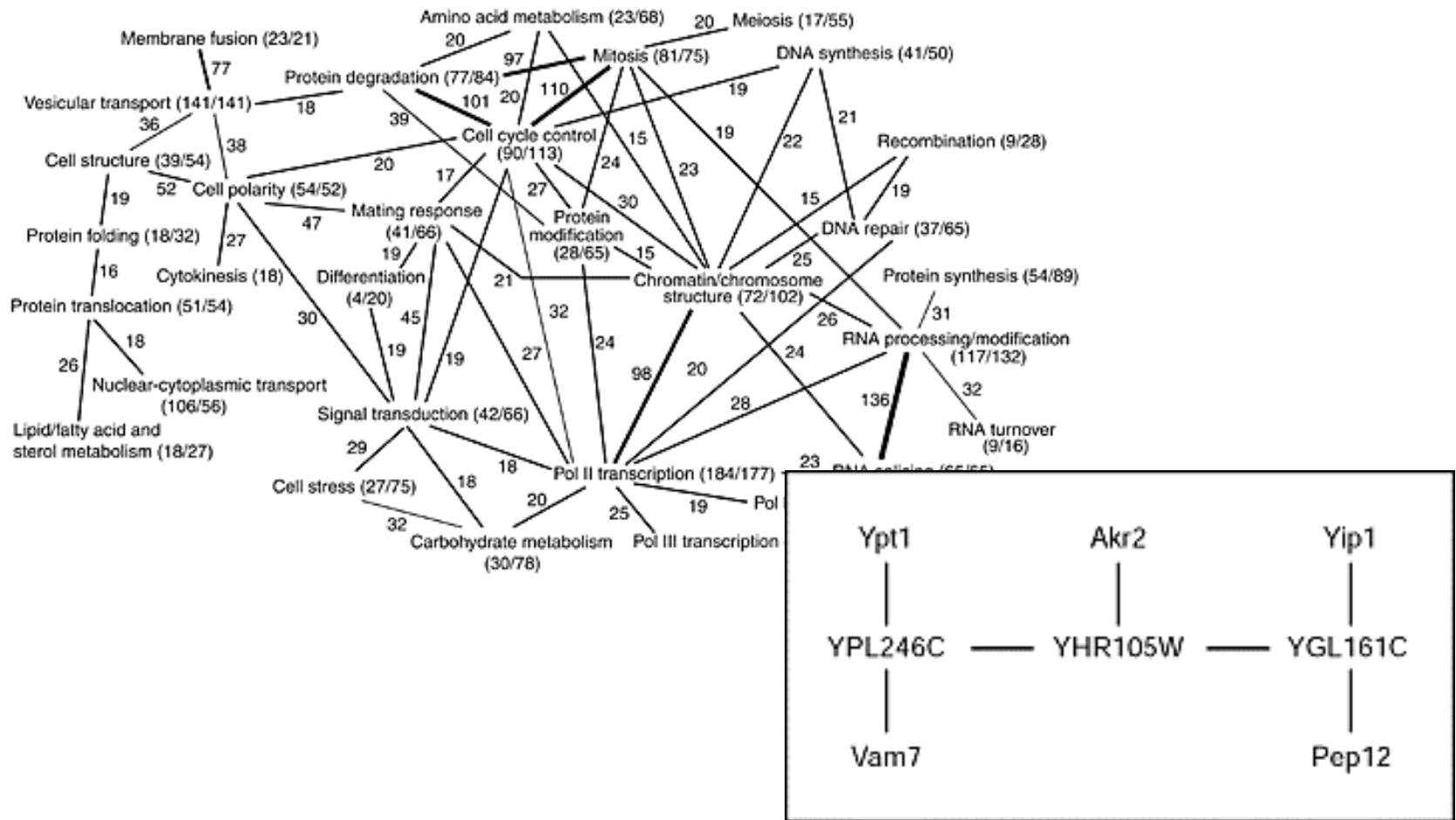


Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature* **411**, 41-42.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. (2000). The large scale organisation of metabolic networks. *Nature* **407**, 651-653.

Schwikowski, B., Uetz, P. & Fields, S. (2002). A network of protein-protein interactions in yeast. *Nature Biotech* **18**, 1257-1261.

# Function Prediction by “Context”



• Vazquez, A., Flammini, A., Maritan, A. and Vespignani, A. (2003) Global protein function prediction from protein-protein interaction networks. *Nat Biotechnol*, **21**, 697-700.

• Samanta, M.P. and Liang, S. (2003) Predicting protein functions from redundancies in large-scale protein interaction networks. *Proc Natl Acad Sci U S A*, **100**, 12579-12583.



# Public repositories of EXPERIMENTAL and PREDICTED protein interactions

The screenshot displays two web browser windows. The left window shows the DIP (Database of Interacting Proteins) interface with a search for 'mraZ'. The right window shows the STRING database interface with a search for 'functional protein association networks'. A large inset box highlights the search results for 'mraZ' in the STRING database.

**Your Input:**  
 MRAZ Protein mraZ (152 aa)

**Predicted Functional Associations:**

		Neighborhood	Gene Fusion	Cooccurrence	Expression Databases	Textmining [Homology]	Score
MRAW	S-acetylmethyltransferase mraW (EC 2.1.1.-) (313 aa)	●				●	0.973
PBPB	Peptidoglycan synthetase ftsI precursor (EC 2.4.1.129) (Peptidoglycan [...])	●				●	0.945
MURX	Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-M [...])	●				●	0.922
MURE	UDP-N-acetylmuramoylalananyl-D-glutamate--2,6-diaminopimelate ligase (EC [...])	●				●	0.921
MRAR	Cell division protein ftsL (121 aa)	●				●	0.918
MRA	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase (EC 6.3.2.1 [...])	●				●	0.913
MURD	UDP-N-acetylmuramoylalanine--D-glutamate ligase (EC 6.3.2.9) (UDP-N-ac [...])	●				●	0.905
MURG	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl [...]	●				●	0.901
FTSW	Cell division protein ftsW (414 aa)	●				●	0.868
MURC	UDP-N-acetylmuramate--L-alanine ligase (EC 6.3.2.8) (UDP-N-acetylmuram [...])	●				●	0.866

**Views:**

- Neighborhood
- Fusion
- Occurrence
- Coexpression
- Experiments
- Databases
- Textmining
- Summary Network

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

von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P. and Snel, B. (2003) STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res*, **31**, 258-261.