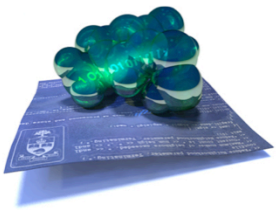
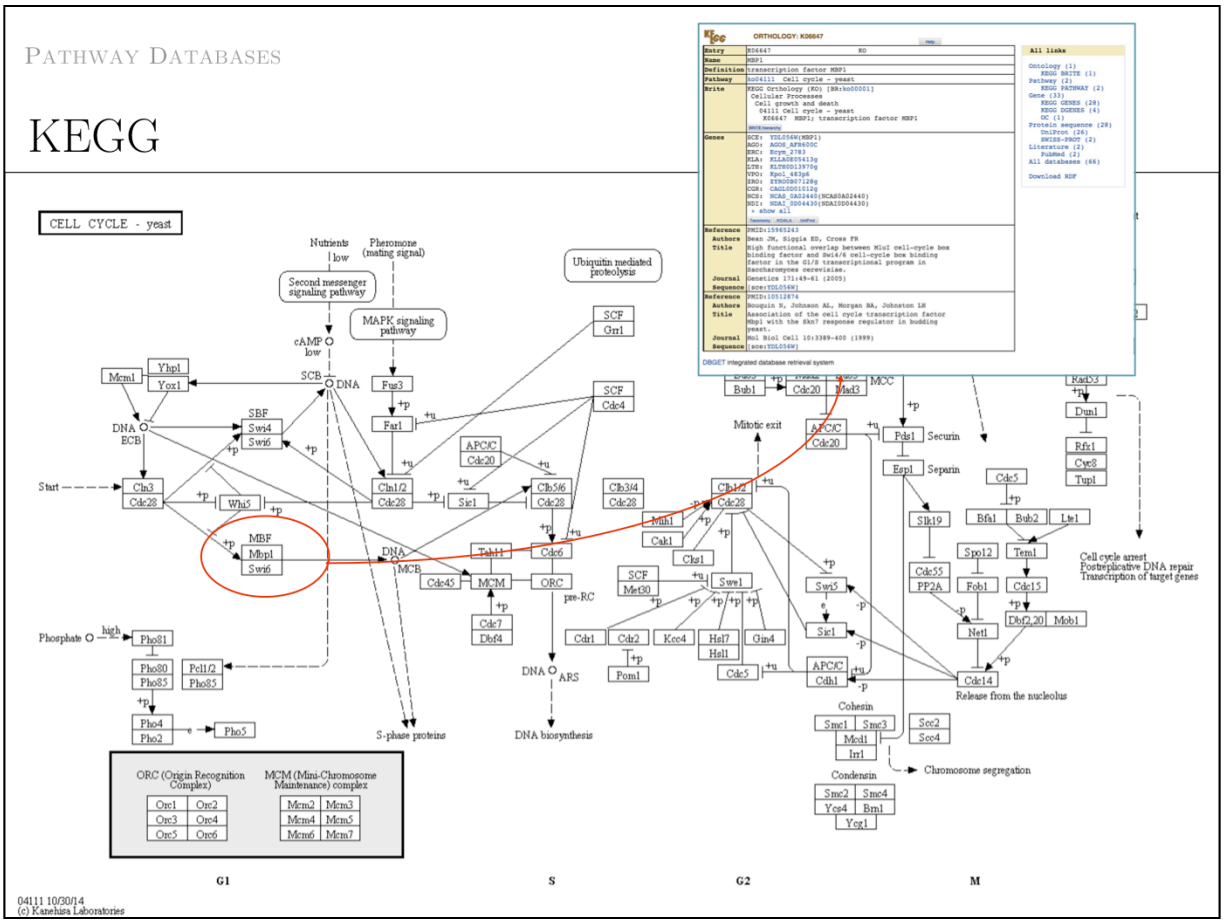
A
Bioinformatics
Course

Concepts of PPI Bioinformatics



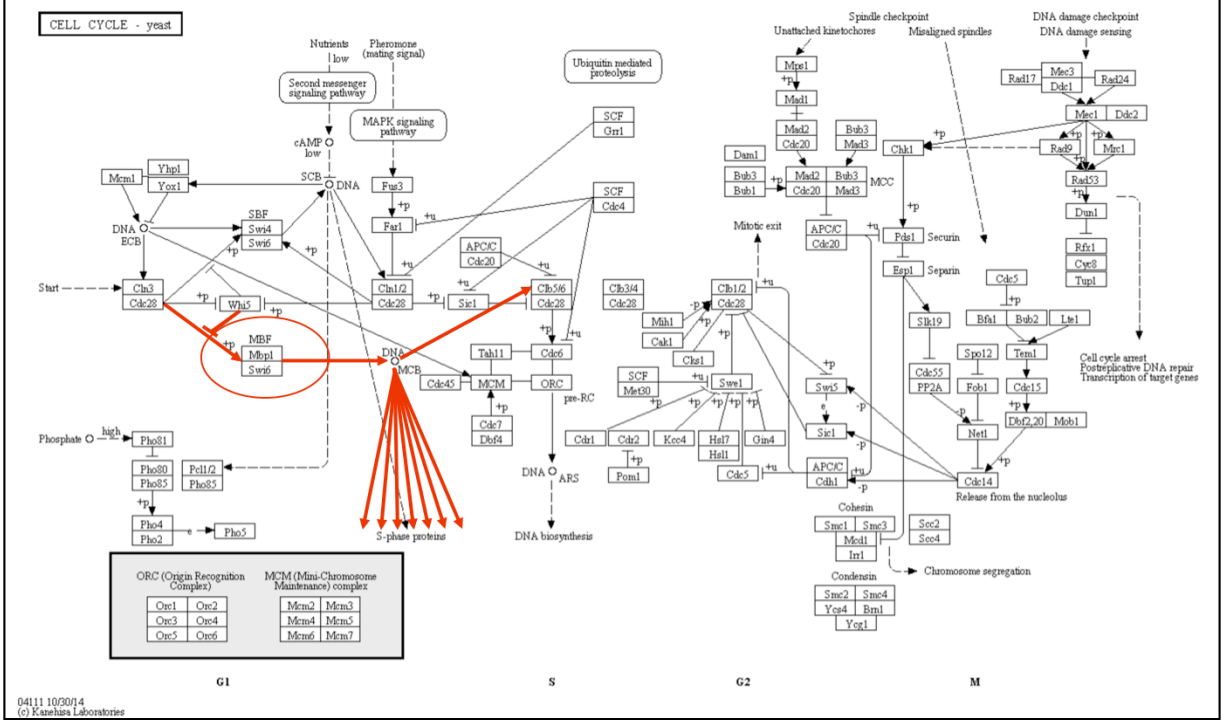
Boris Steipe

department of biochemistry – department of molecular genetics
University of Toronto



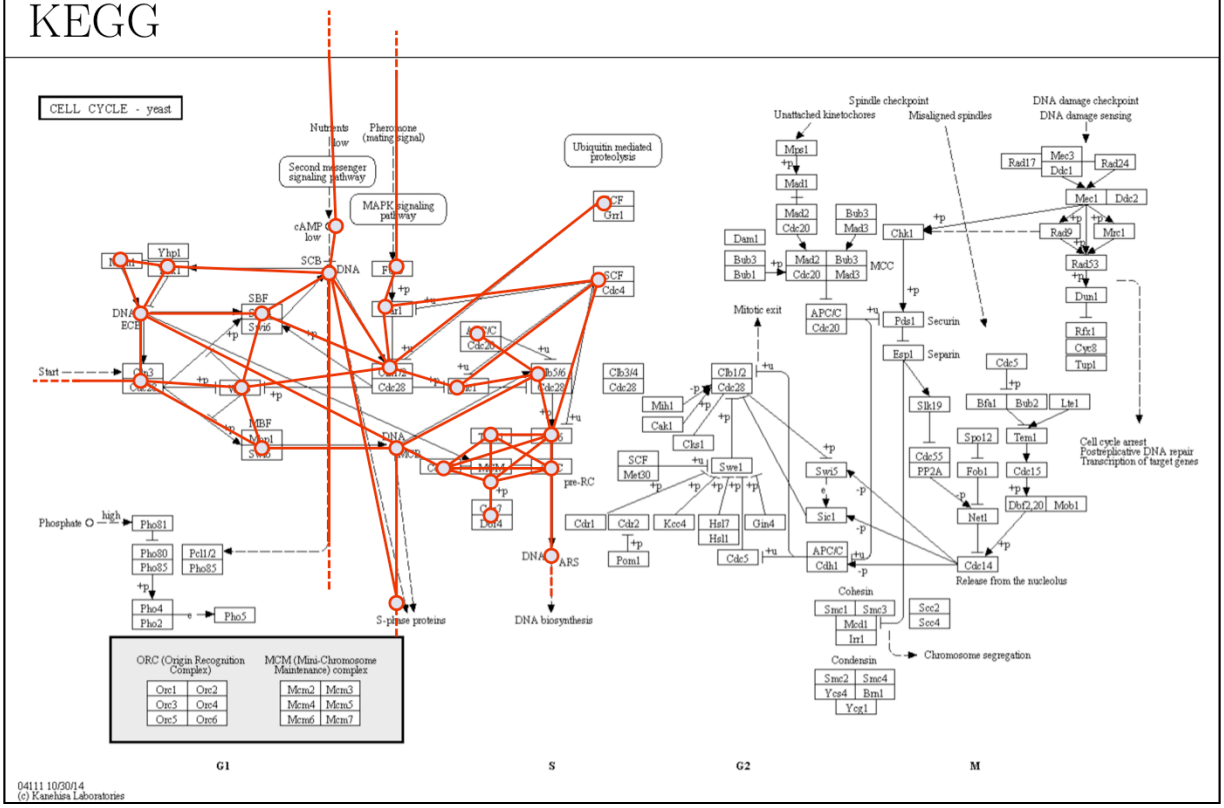
Individual biomolecules have properties that can be stored in databases – such as sequences, anotations, structures and function.

KEGG



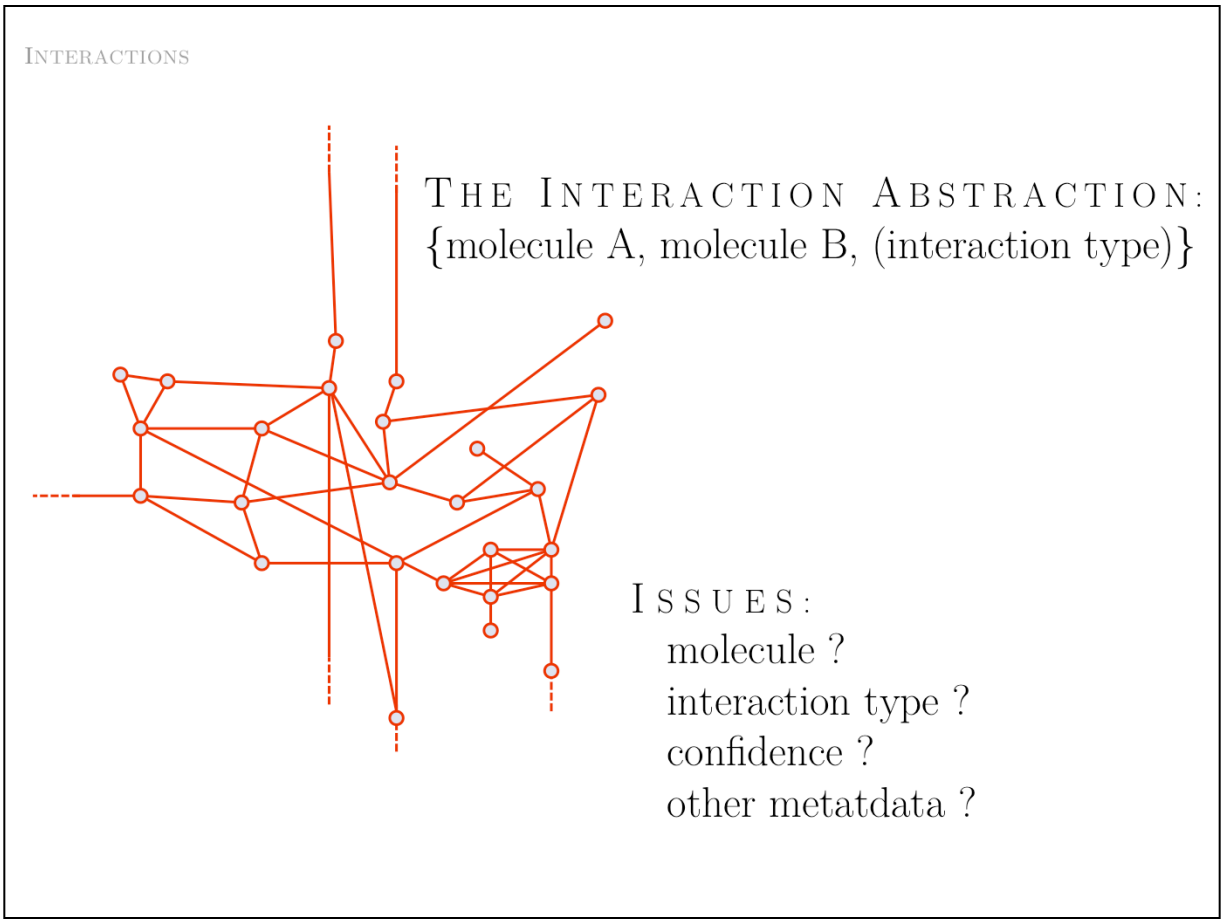
But the **roles** of biomolecules require them to work in *pathways* – as collections of functions – or in *systems* of collaborating genes.

KEGG



This means: to analyze biomolecular roles and functions we need to consider their interactions. If we think of entities and relationships, we need to focus on the relationships, not the entitites to describe and analyze function.

The large-scale integration of biomoleculaes and functional systems into the whole of metabolism, developmental regulation, or the coordinated response to external stimuli can be described as networks of structural and functional interactions.



The abstraction we use to describe these interactions is a tuple of three components. Two *molecules*, usually proteins, but these could also include metabolites and certainly DNA and RNA, and an *interaction*. Such tuples are fundamental to graphs.

In practice, to capture biological reality, a number of issues need to be addressed. These include defining the interacting molecules (these could be genes, but these could also be restricted to particular transcripts, post-translational modifications, or cellular loactions). As well, "interactions" can have many different types: activating and repressing, modifying or modulating, and many more directed and undirected types. Finally, interactions are determined with particular experiments and the experiments are limited in accuracy and precision. Other metadata may need to be associated with the interaction as well.

A major effort is underway to experimentally describe all physiological protein-protein interactions (PPI) in the cell.

The Experiments:

Biochemical:

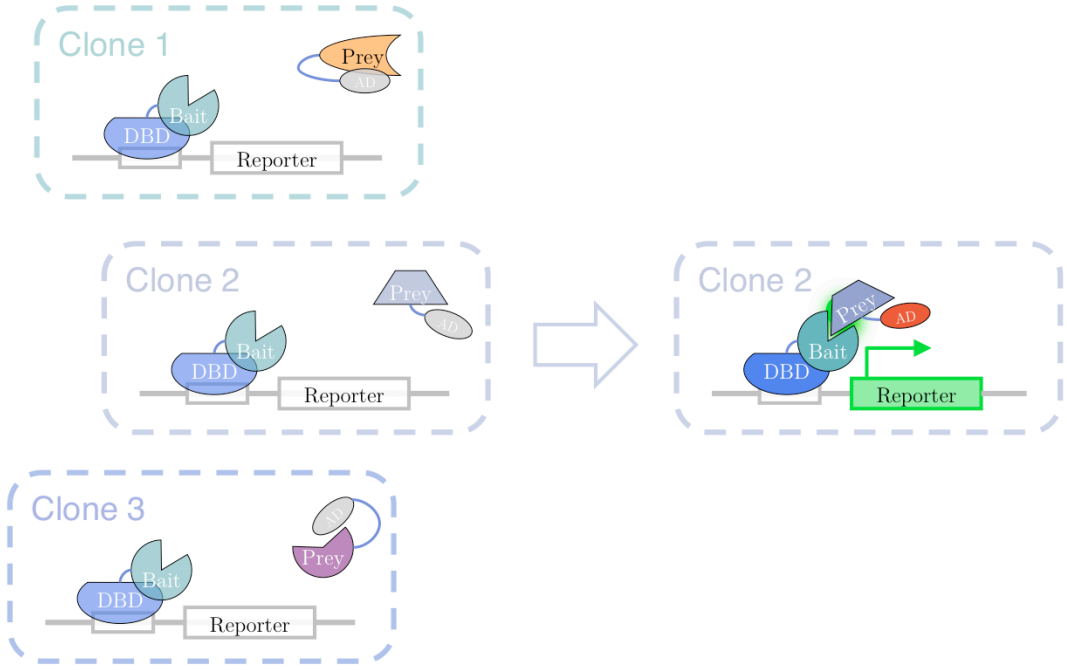
- Yeast 2-Hybrid
- Tap-tag MS/MS
- Split ubiquitin and other PAC
- Co-IP
- Crosslinking (chemical and in vivo)
- Phage display

Biophysical:

- FRET
- SPR
- Correlation spectroscopy
- ITC
- Analytical ultracentrifugation
- Crystallography / NMR

There are many sources of protein-protein interaction data: besides biochemical and biophysical methods, data are contributed from predictions based on homologous sequences, and so called “genetic interactions” or epistatic effects.

YEAST TWO-HYBRID :



The first large scale interaction datasets came from yeast two-hybrid experiments.

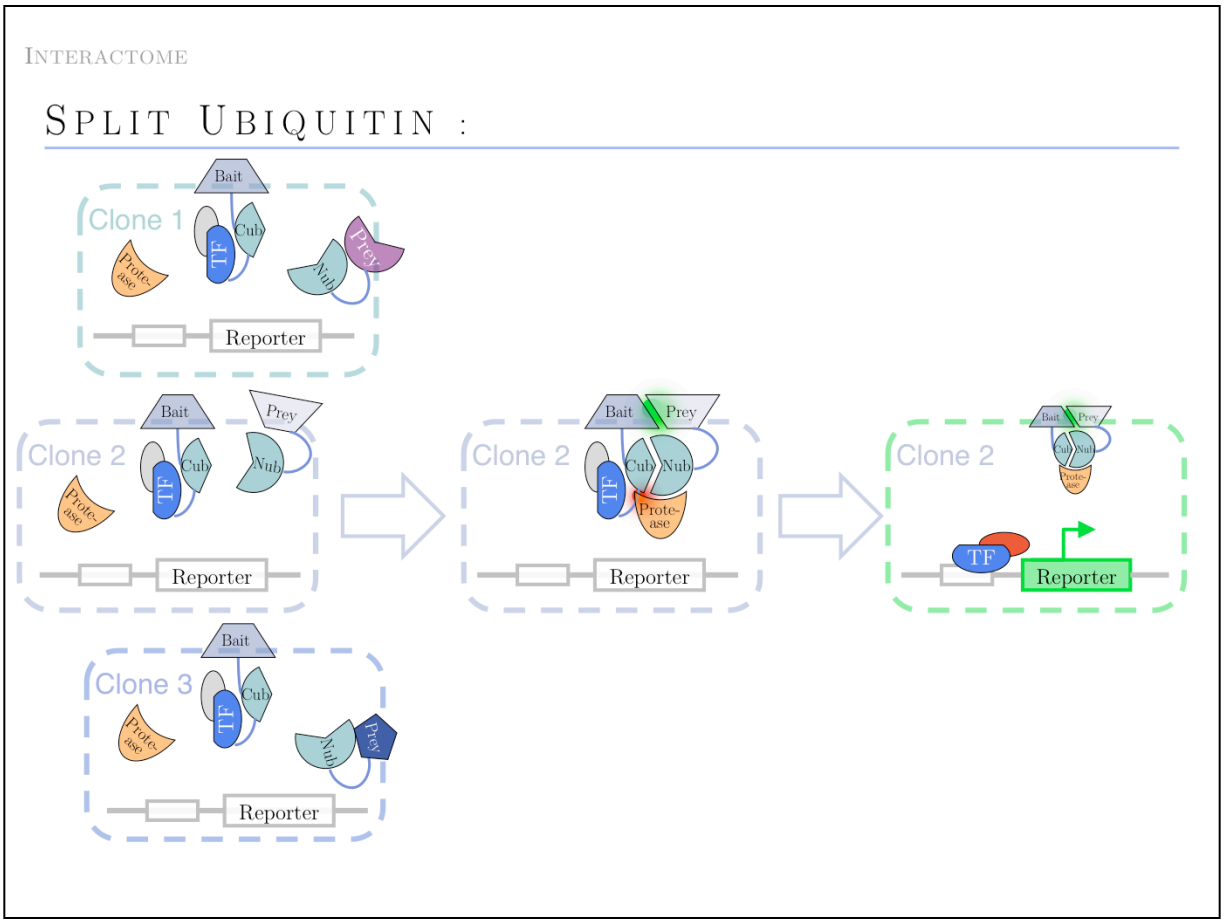
A “Bait” protein is fused genetically to a DNA-binding domain (DBD). Cells with this construct are then crossed with a library of cells with “Prey” proteins (Clones 1-3), that are genetically fused with a RNA-polymerase activation domain (AD).

Transcription of a reporter (or survival) gene is induced in clones in which the bait protein can bind to the prey protein.

Although this experiment can be done in a high-throughput mode, there are a number of situations in which errors can occur.

False negatives (can’t detect a physiological interaction) are expected when the fusions interfere with the interaction, and when the interaction can’t take place in the nucleus.

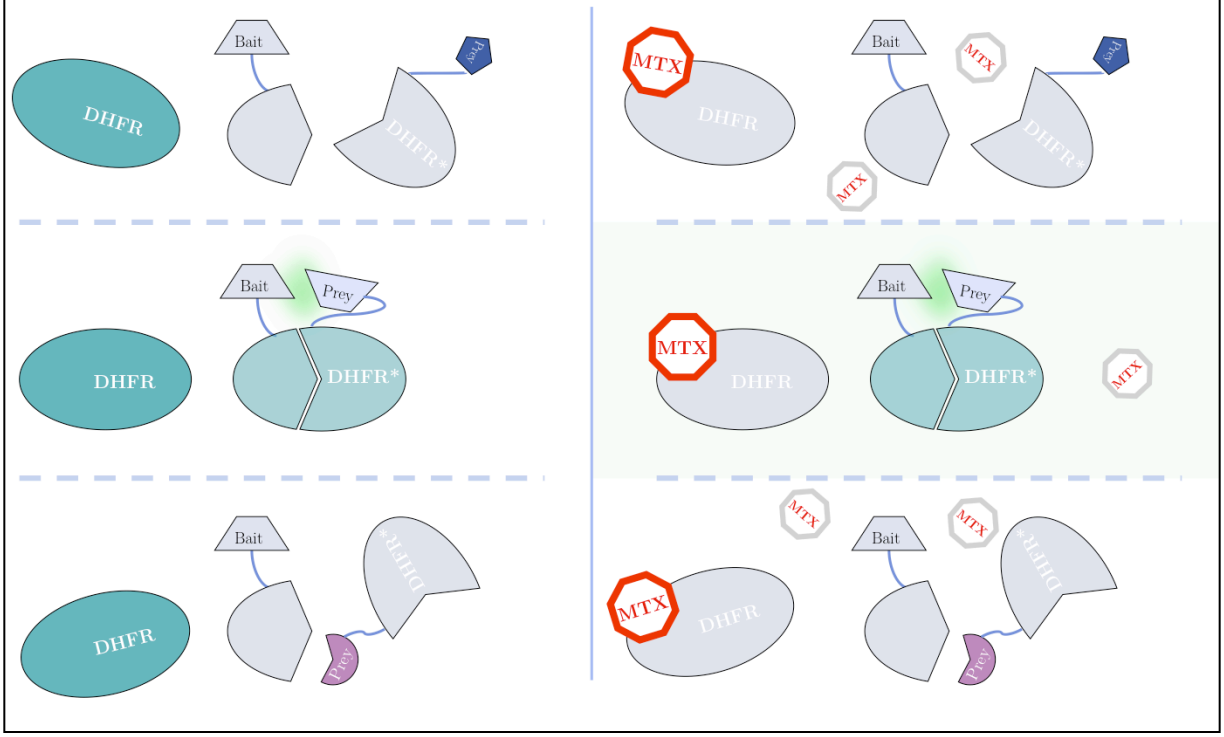
False positives (detected interaction is not physiological) can occur if the prey can bind to the DBD, or the bait to the AD, if the bait can activate polymerase, if the interaction is forced by the unphysiologically high recombinant expression, or if prey and bait bind to a third molecule.



In split-Ubiquitin assays, the reconstitution of ubiquitin, induced by an interacting bait/prey pair, provides a substrate for a ubiquitin protease, which cleaves off a genetically fused reporter domain. The reporter protein commonly used is a transcription factor, which drives expression of a reporter gene (or survival factor).

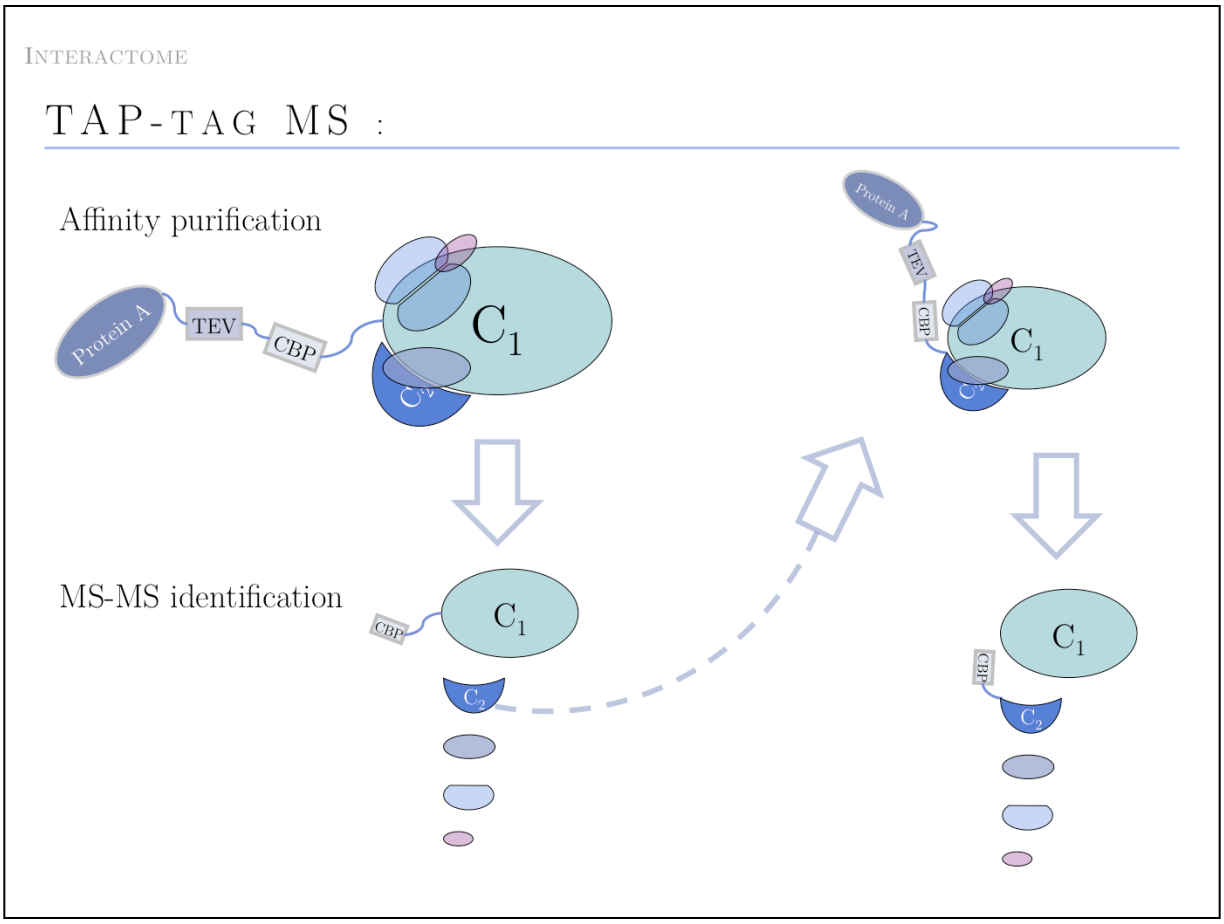
This method is not restricted to the nucleus and has been used successfully for detecting membrane protein interactions.

PCA (Protein-fragment Complementation Assay):



In protein-fragment complementation assays, a protein is also genetically split into two parts that can be induced to reassemble into a functioning protein. In the DHFR example, the proximity of prey and bait is sufficient to reconstitute DHFR activity, a tight interaction is not required.

The reconstituted DHFR* is resistant to the anti-metabolite methotrexate, which inhibts wild-type DHFR. Therefore only cells with a reconstituted DHFR* – revealing a bait/prey interaction – survive.



AP-MS (affinity purification mass spectrometry) is often performed as a TAP-tag experiment (Tandem Affinity Purification – tag). The original version of the system used a tag of calmodulin binding peptide – TEV protease recognition site – protein A, to be purified on IgG resp. calmodulin columns, and eluted with TEV protease and EGTA, respectively.

Once the protein complex is purified, its constituents are identifed by mass-spectrometry.

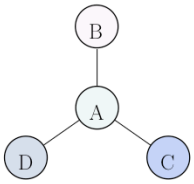
A validation experiment can be undertaken in which one of the newly identifed proteins is tagged, and expected also to purify the whole complex.

Note that the washing steps essentially place the complex into infinite dilution, and this may bias for complex components with slow off-rates. Therfore the method is not considered quantitative.

Similar method: co-fractionation MS.

Binary vs. N-ary Data:

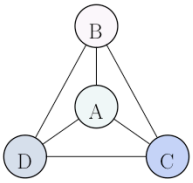
Interpreting complex membership as spoke–, matrix– or complex models for storage of binary data in interaction databases.



Spoke model:

A–B
A–C
A–D

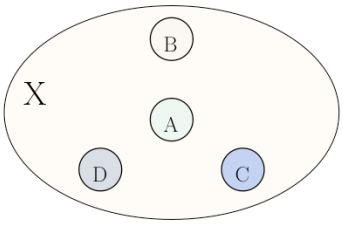
(underestimates)



Matrix model:

A–B B–C C–D
A–C B–D
A–D

(overerestimates)



Complex model:

X–A
X–B
X–C
X–D

(X is not a protein)

Proteins that are co-purified in a complex need to be interpreted as binary interactions to store them in databases according to the commmon binary interaction abstraction. This can be done as a spoke- or matrix- model, or by recording a complex as an abstract entity (X in our example).

INTEROLOGS

Biological interaction networks are conserved at the module level

Zinman GE, Zhong S, Bar-Joseph Z (2011) BMC Systems Biology, 5:134

Background: Orthologous genes are highly conserved between closely related species and biological systems often utilize the same genes across different organisms. However, while sequence similarity often implies functional similarity, interaction data is not well conserved even for proteins with high sequence similarity. Several recent studies comparing high throughput data including expression, protein-protein, protein-DNA, and genetic interactions between close species show conservation at a much lower rate than expected.

Results: In this work we collected comprehensive high-throughput interaction datasets for four model organisms (*S. cerevisiae*, *S. pombe*, *C. elegans*, and *D. melanogaster*) and carried out systematic analyses in order to explain the apparent lower conservation of interaction data when compared to the conservation of sequence data. We first showed that several previously proposed hypotheses only provide a limited explanation for such lower conservation rates. We combined all interaction evidences into an integrated network for each species and identified functional modules from these integrated networks. We then demonstrate that interactions that are part of functional modules are conserved at much higher rates than previous reports in the literature, while interactions that connect between distinct functional modules are conserved at lower rates.

Conclusions: We show that conservation is maintained between species, but mainly at the module level. Our results indicate that interactions within modules are much more likely to be conserved than interactions between proteins in different modules. This provides a network based explanation to the observed conservation rates that can also help explain why so many biological processes are well conserved despite the lower levels of conservation for the interactions of proteins participating in these processes.

Accompanying website: http://www.sb.cs.cmu.edu/CrossSP

http://steipe.biochemistry.utoronto.ca/abc

B O R I S . S T E I P E @ U T O R O N T O . C A

DEPARTMENT OF BIOCHEMISTRY & DEPARTMENT OF MOLECULAR GENETICS
UNIVERSITY OF TORONTO, CANADA