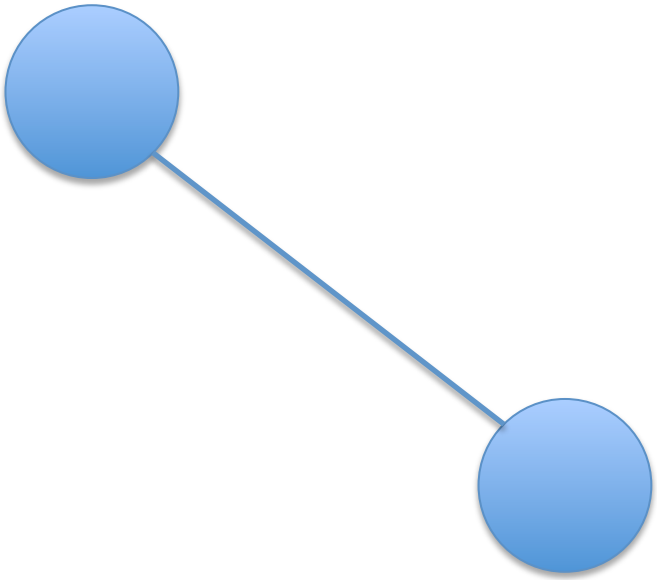
Protein-Protein Interaction Network

Lecture 1

Outline

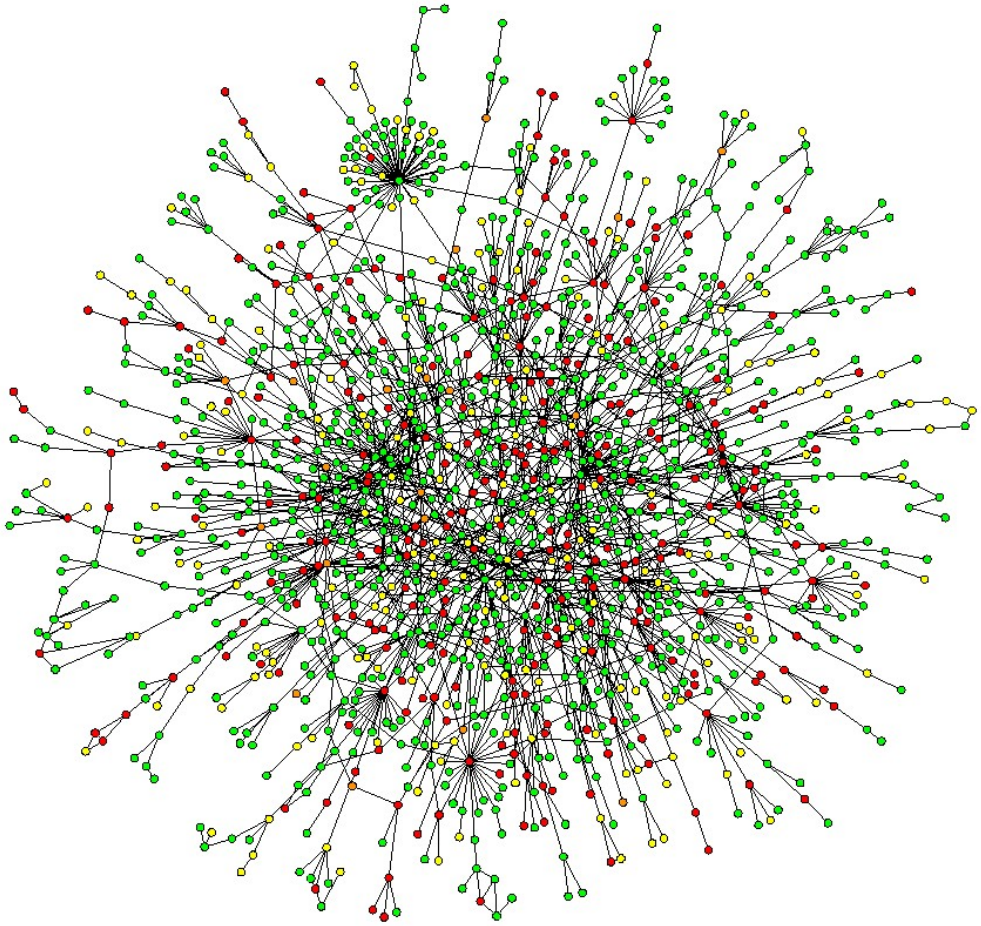
- Protein-Protein Interaction Model
- How to get PPI
  - Experimental methods (methods, results, assessing and filtering)
  - Bioinformatic methods
- PPI databases
- network properties
- Analysis method
- Integration with other omic data

Graph Model



Vertex
Edge

Yeast protein interaction network

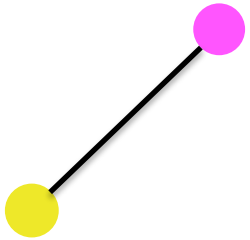


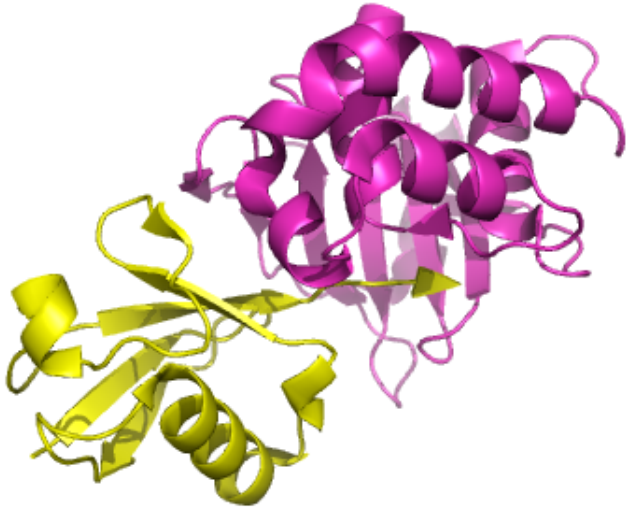
What kind of interactions?

- Protein Physical Interactions
  - Protein-protein binding
  - Enzyme and its substrates
  - Enzyme and its inhibitor
  - Protein Chaperon
  - Protein complexes

Protein Binding

- L-protein and ubiquitin

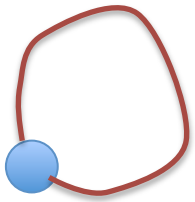


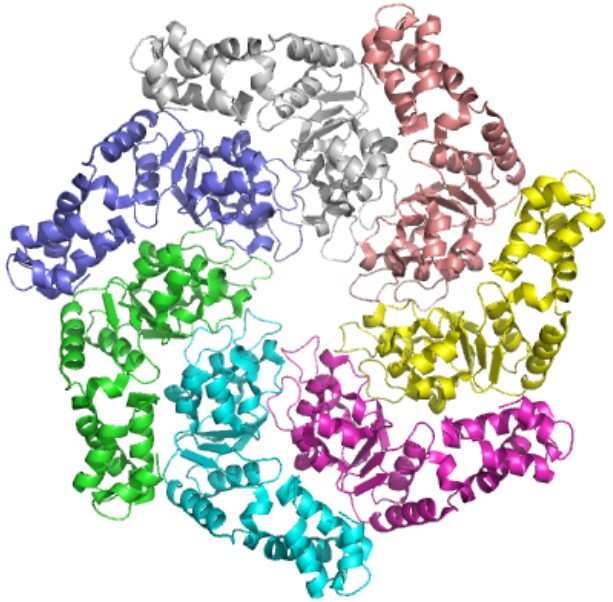


PDB: 3PRP

Protein Binding

- NtrC1 ATPase domains form a Heptamer

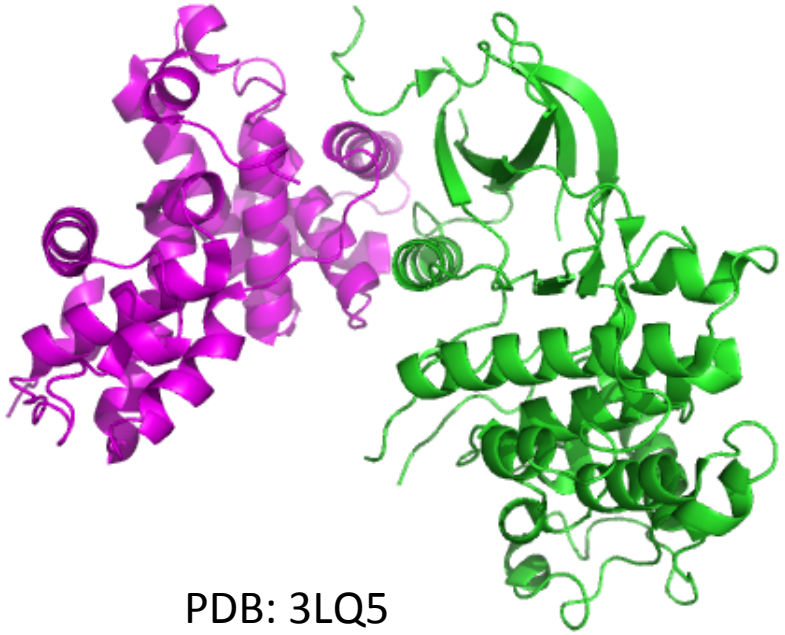




3M0E

Enzyme and its substrate

- Cell division protein kinase 9 and Cyclin-T1
- Trigger Mcl-1 Down-Regulation and Apoptotic Cell Death in Neuroblastoma Cells



PDB: 3LQ5

Enzyme and its inhibitor

- Xylanase is a class of enzymes which degrade the linear polysaccharide beta-1,4-xylan into xylose, thus breaking down hemicellulose, one of the major components of plant cell walls.

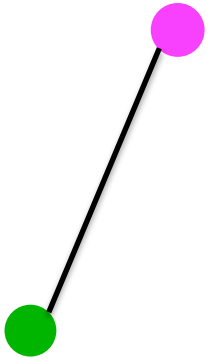
Xylanase (Enzyme)

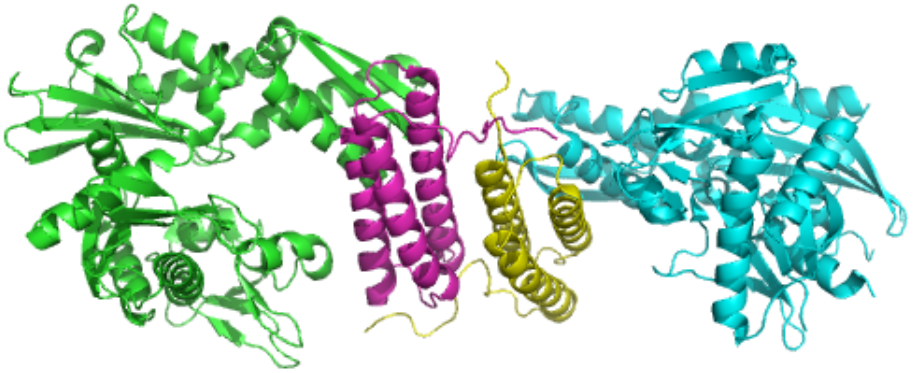
TAXI (Inhibitor)

PDB: 2B42

Protein Chaperone

- Complex between the BAG5 BD5 and Hsp70 NBD

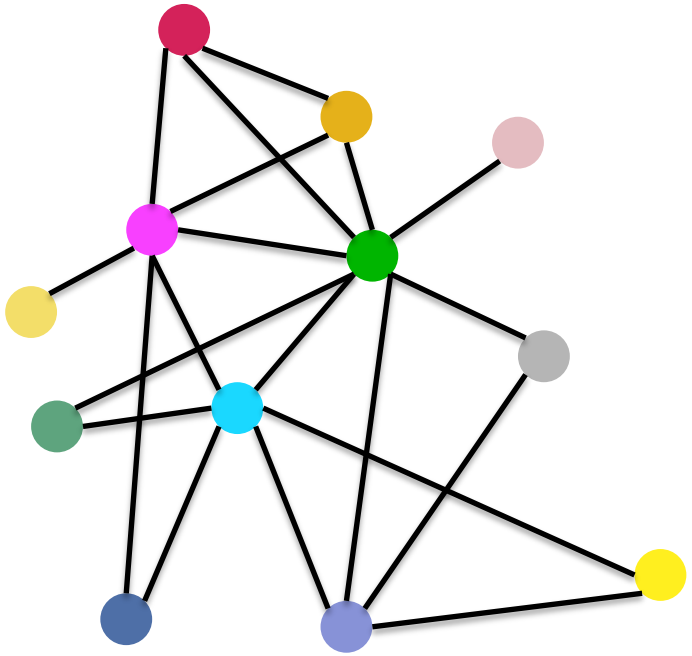


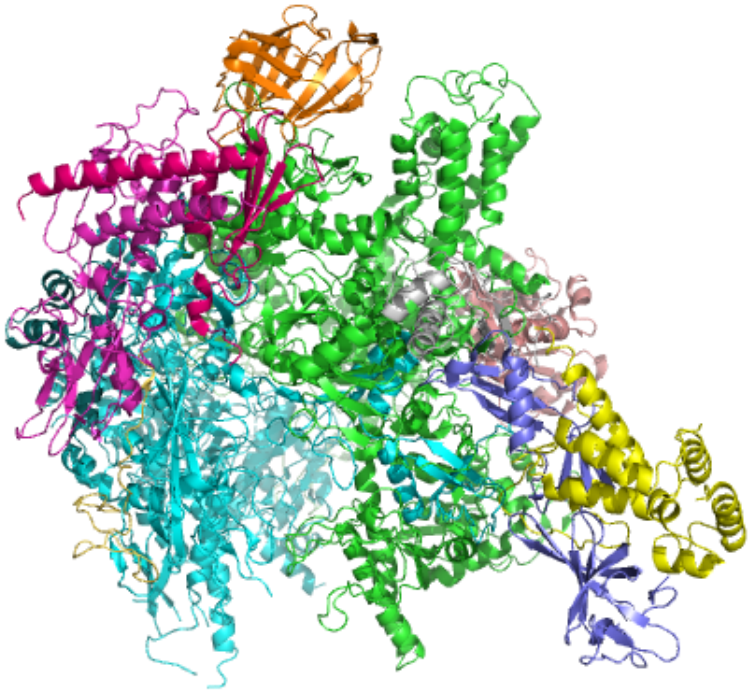


PDB: 3A8Y

Protein Complex

- 12-subunit RNA Polymerase II

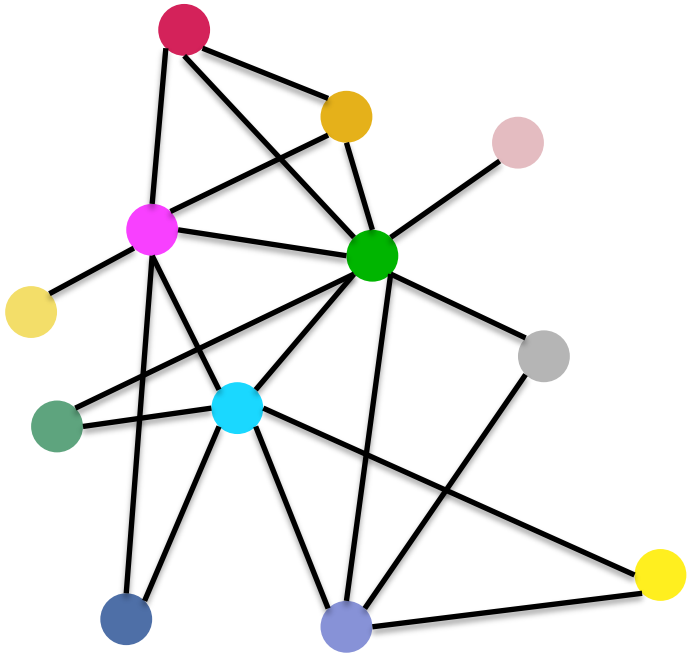




PDB: 2B8K

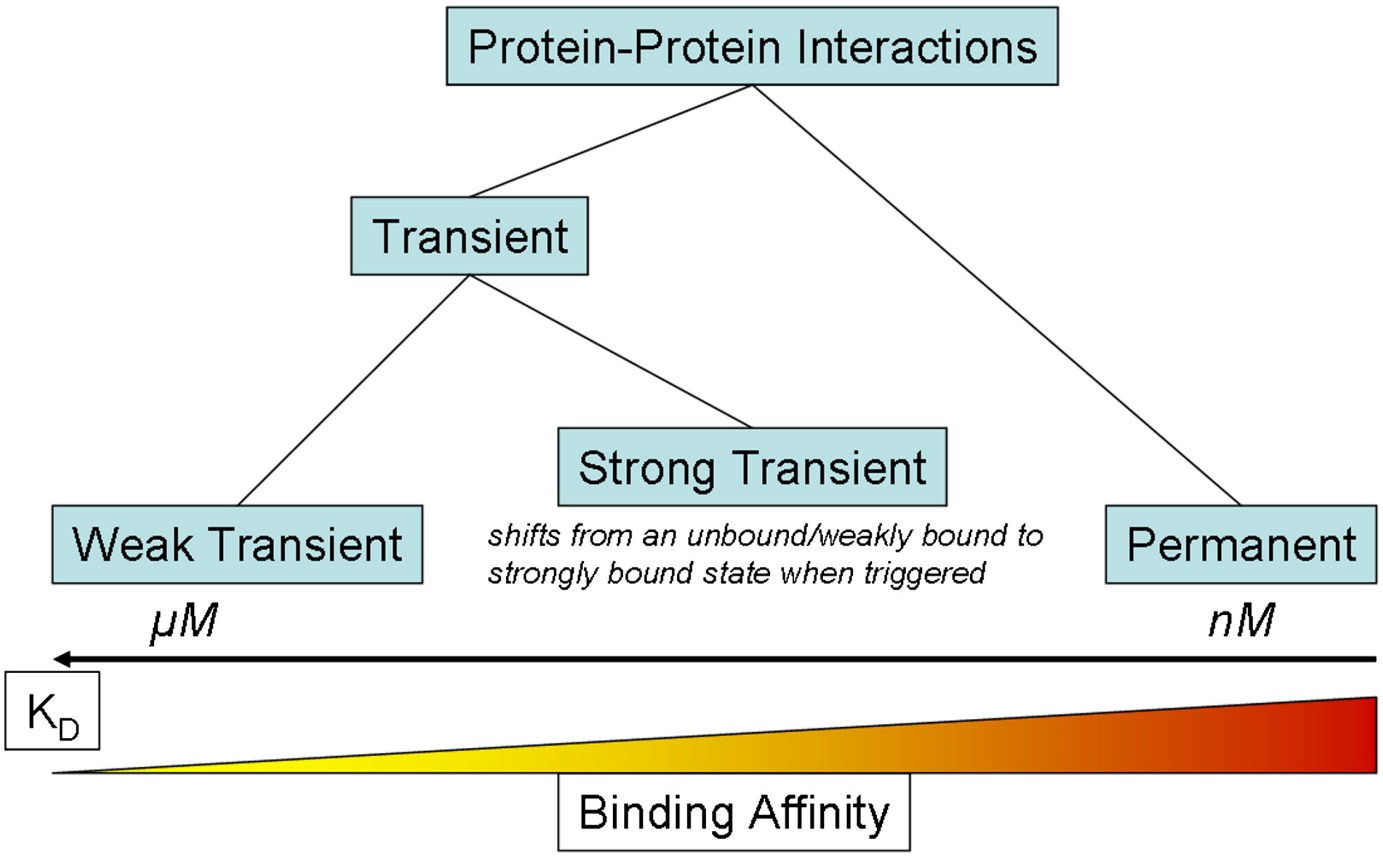
Protein Complex

- What is the connection density for this graph?



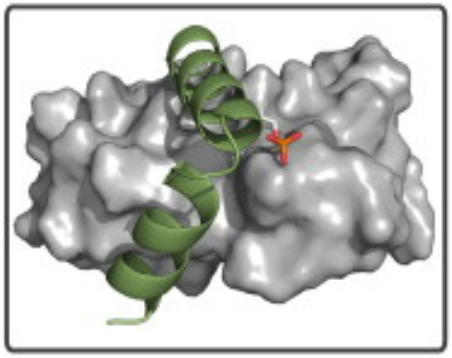
$$Q = \frac{|E|}{V(V-1)/2}$$

Permanent or Transient interactions

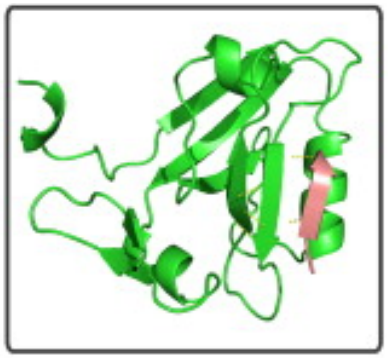


Perkins *et al.* Structure (2010)

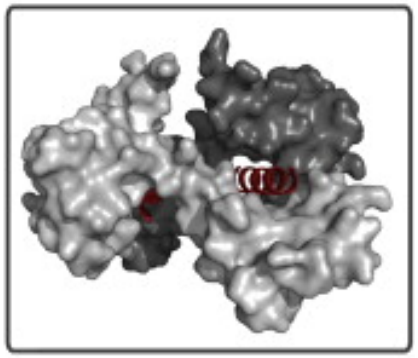
Permanent or Transient interactions



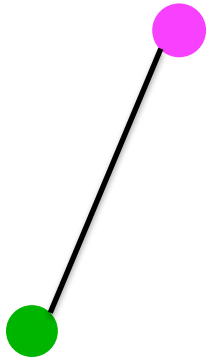
A KIX domain of CBP with KID peptide of CREB



B PSD-95 PDZ domain with its peptide



C Calcineurin - Calmodulin complex



- Difficult to measure the transient interactions.
- How to distinguish permanent and transient interactions in PPI network?

Perkins *et al.* Structure (2010)

What kind of information PPI network cannot provide?

- Protein binding affinity? No
- Network topology? Yes
- Protein binding interface? No
- Protein function? We will try

PPI networks for entire genomes

- The potential number of interactions is huge, and the number of real interactions is probably very large.
  - ~16 000–26 000 different interaction pairs in the yeast. Grigoriev Nucleic acid Research (2003)
  - ~600,000-250,000,000 interaction pairs in human genome.
- However, the current status to the knowledge of those interactions is still poor; only a small portion of those protein interaction pairs have been discovered.
- The large amount of interaction pairs is also a challenge to study them. The “network” is a suitable tool to study on the PPI data.

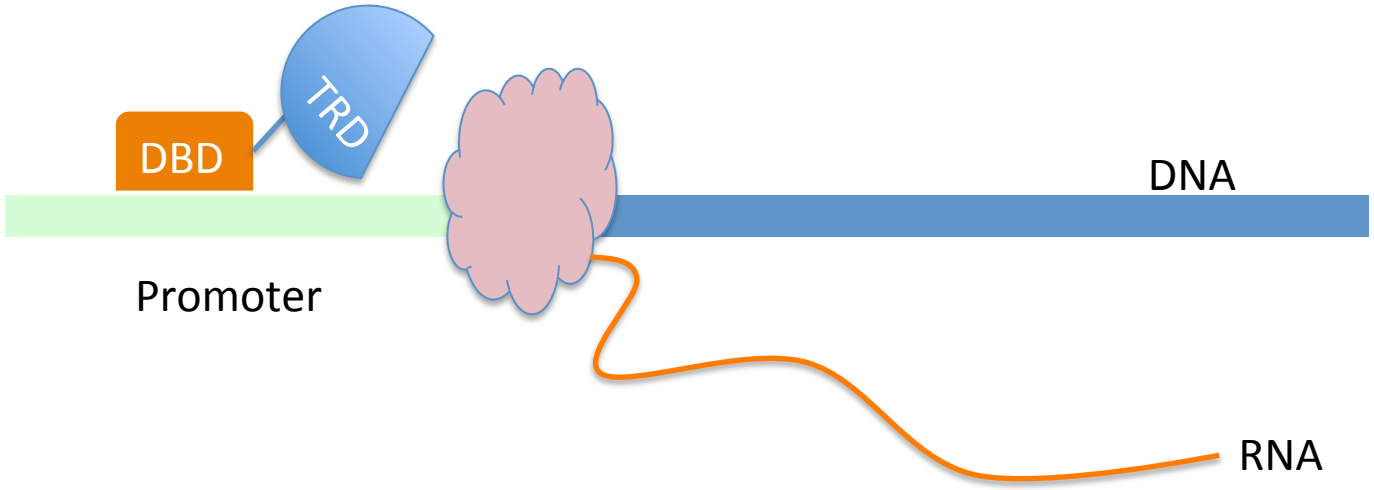
Outline

- Protein-Protein Interaction Model
- How to get a PPI network
  - Experimental methods: Y2H, MS etc.
  - Bioinformatic methods
- PPI databases and network properties
- Analysis method
- Integration with other omic data

Experimental methods

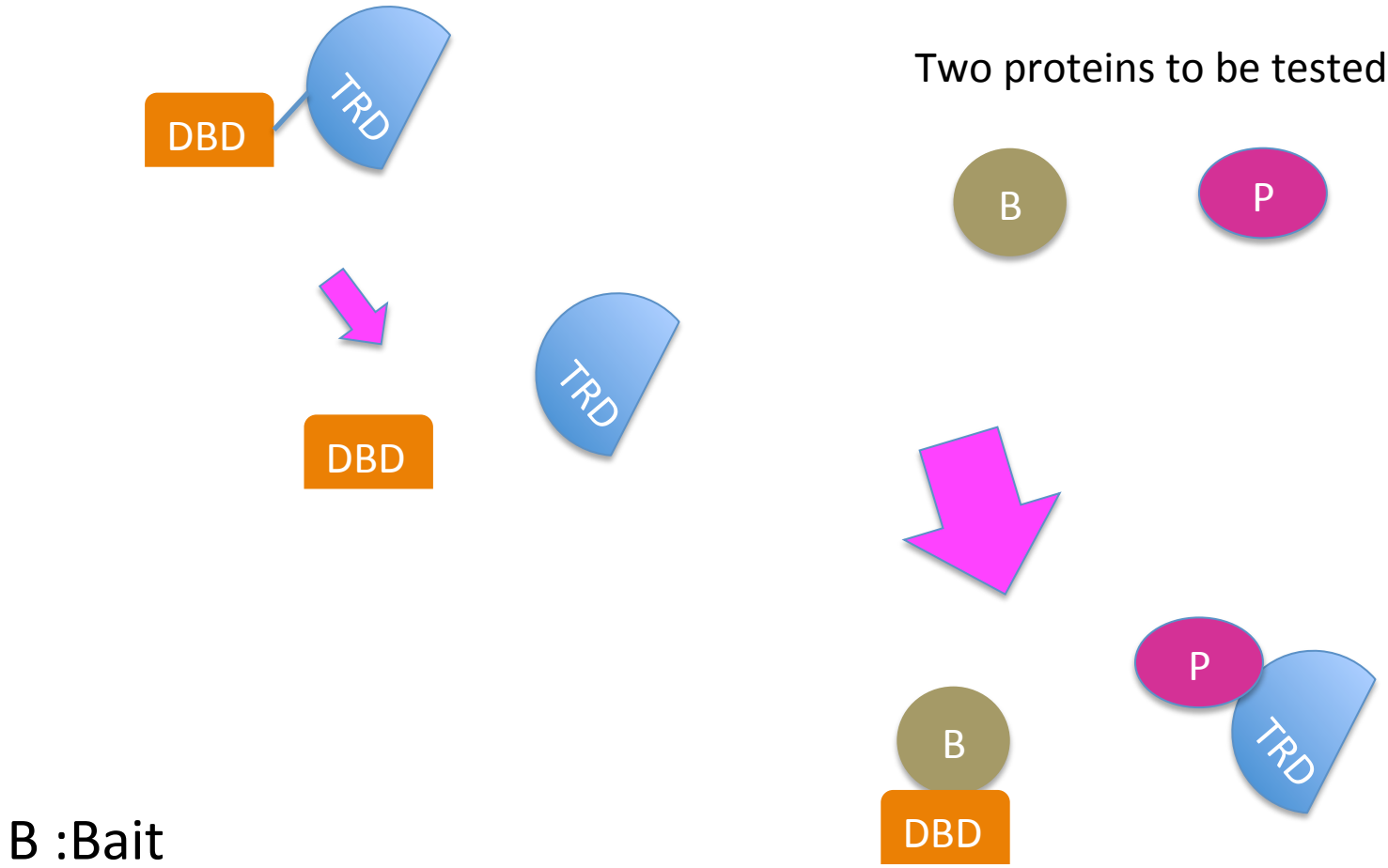
- **Co-immunoprecipitation** is considered to be the gold standard assay for protein–protein interactions, especially when it is performed with endogenous (not overexpressed and not tagged) proteins.
- **Pull-down assays** are a common variation of immunoprecipitation and are used identically, although this approach is more amenable to an initial screen for interacting proteins.
- **Chemical cross-linking** is often used to "fix" protein interactions in place before trying to isolate/identify interacting proteins.
- **Yeast two-hybrid assay**
- **Tandem Affinity purification**
- **Protein microarray**
- **Phage display**

Yeast Two-hybrid Assay



DBD :DNA binding domain
TAD: Transcriptional Activation domain

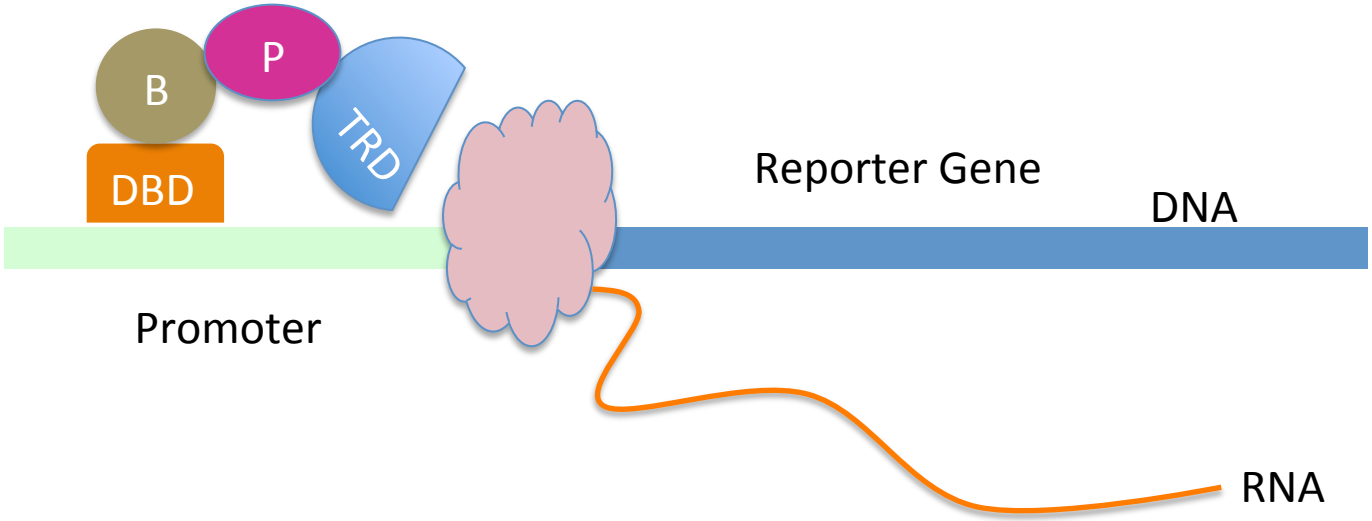
Yeast Two-hybrid Assay



B :Bait

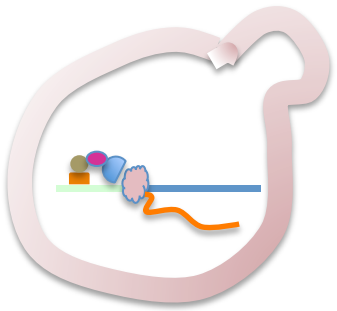
P: Prey

Yeast Two-hybrid Assay

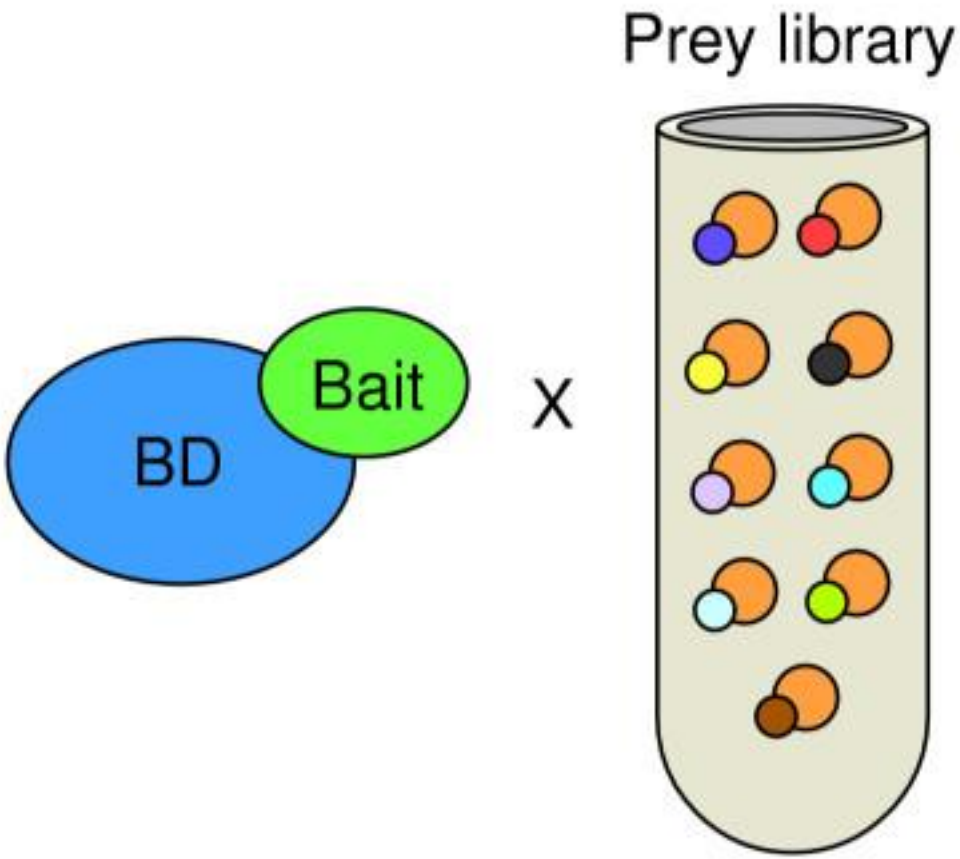


Transcription factor: Gal4

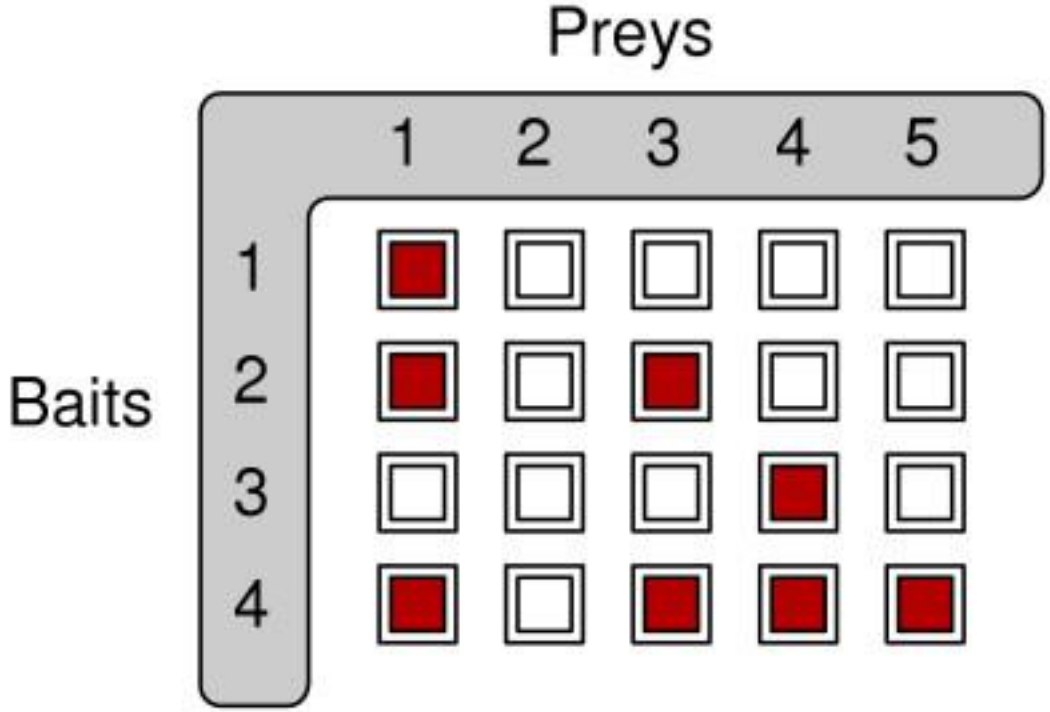
Reporter gene: LacZ



Yeast Two-hybrid Assay

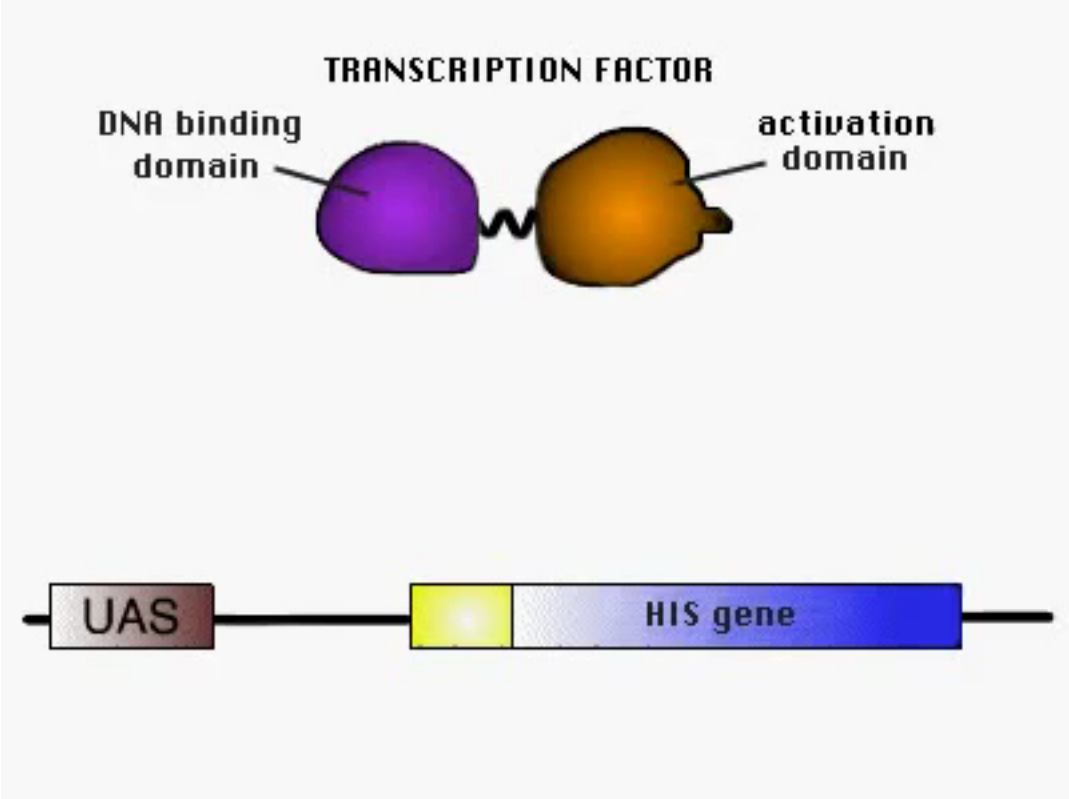


Yeast Two-hybrid Assay



What does this matrix is?

Yeast 2-hybrid Assay Video



http://www.youtube.com/watch?v=vk3I4xldC7o

Yeast 2-hybrid Assay

- Pros
  - Easy/fast
  - No purification required
  - *In vivo* conditions
  - Can be adapted for high throughput screens
  - Can detect transient interactions

Yeast 2-hybrid Assay

- Cons
  - prone to false negatives because
    - protein doesn't fold,
    - protein doesn't localize to nucleus,
    - interference from endogenous protein,
    - fusion protein doesn't interact like native protein,
    - fusion may be toxic to cell
  - prone to false positives
    - auto-activation
    - indirect interactions
  - not quantitative
  - no control over post-translational modification
  - only test binary interactions

Yeast 2-hybrid assay for an entire genome

Uetz et al. Nature (2000) 403, 623-627

Two strategies:

1. "array" approach: ~6,000 activation domain hybrid transformants mated to 192 DNA binding domain fusion transformants only 20% of interactions (281) reproducible (many auto-activate), and 3.3 positives per interaction-competent protein
2. "high-throughput screen" approach: 5,345 ORFs cloned separately into DNA-binding and activation domain plasmids (2 reporter genes); DBD fusions pooled and mated to AD fusions; 12 clones per pool sequenced, gave 692 unique interactions (472 seen more than once) 1.8 positives per interaction-competent protein.

More “cons” for Yeast 2-hybrid Assay

- Cloning and transformation inefficiencies
- If baits are pooled, slow-growing cells will lose to faster ones, giving false negatives.
- All vs. all assay contains many implausible interactions -- proteins that aren’t co-localized or expressed at the same time.
- Can only sequence a small fraction of the positive clones.
- High-throughput Y2H screens miss as many as 90% of Y2H interactions observed in focused.

Y2H is still the most popular method to study PPI network

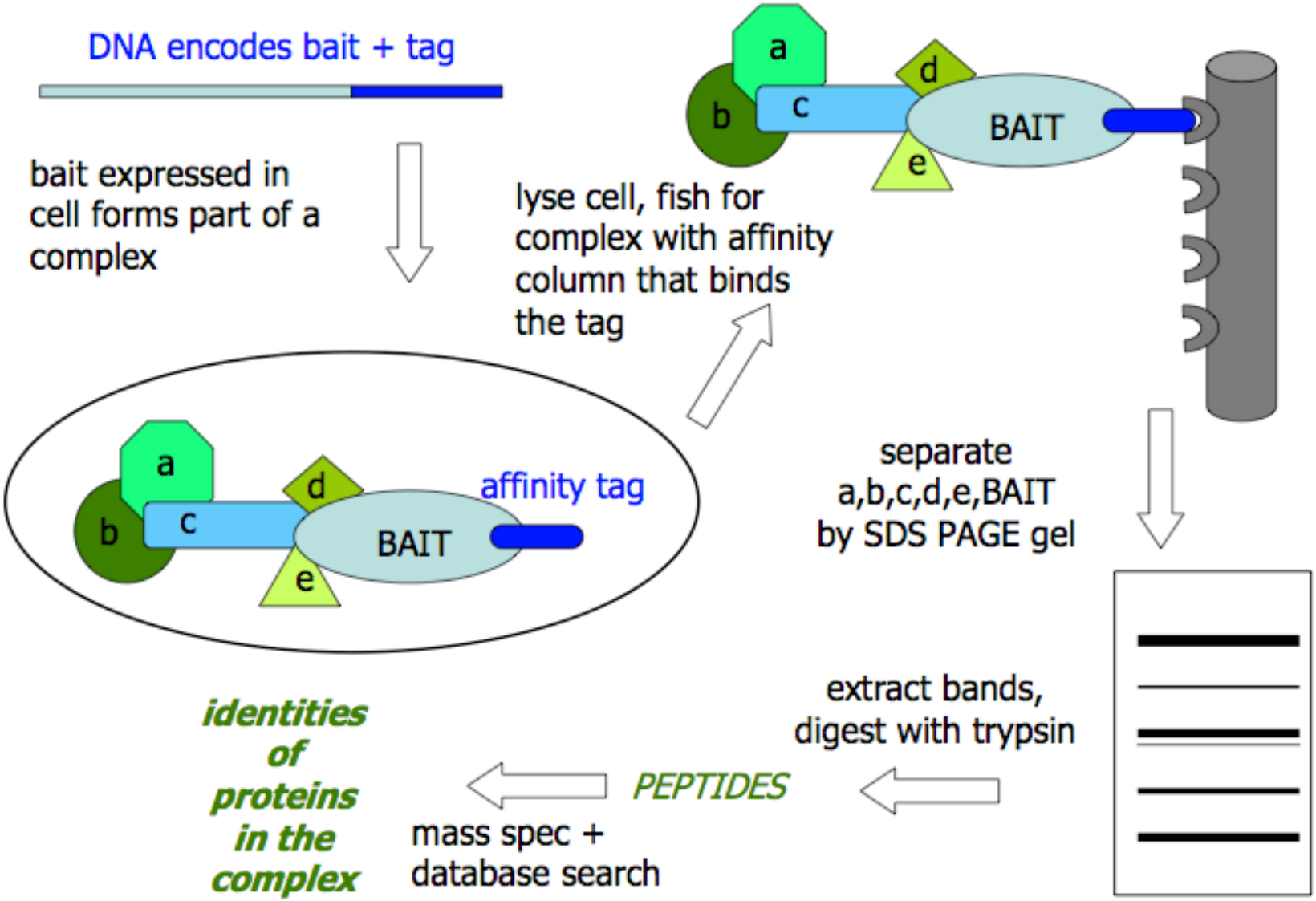
Experimental methods

- **Co-immunoprecipitation** is considered to be the gold standard assay for protein–protein interactions, especially when it is performed with endogenous (not overexpressed and not tagged) proteins.
- **Pull-down assays** are a common variation of immunoprecipitation and are used identically, although this approach is more amenable to an initial screen for interacting proteins.
- **Chemical cross-linking** is often used to "fix" protein interactions in place before trying to isolate/identify interacting proteins.
- **Yeast two-hybrid assay**
- **Tandem Affinity purification (TAP)**
- **Protein microarray**
- **Phage display**

Tandem Affinity Purification (TAP)

- Most proteins interact with several other proteins (estimate 2-10).
- Many proteins in the cell are found in complexes. For some purposes, knowing the identities of the members of the clusters is as useful, or more useful, than knowing the directly interacting partners.
- Tandem Affinity purification (TAP) is a method for characterizing the clusters directly, rather than one interaction at a time.

TAP/MS spectrometry



TAP/MS spectrometry for an entire genome

- Gavin et al. Nature(2002) 415, 141-147;
  - Cellzome 1,167 bait proteins in Yeast genome
  - TAP tag inserted at 3' end of gene; proteins under endogenous promoter 2 rounds of purification
  - 232 distinct complexes with 2 to 83 proteins per complex new cellular role proposed for 344 proteins
  - To assess confidence:

    Repeat the experiment -only 70% reproducible using the same bait
    Use different proteins in the complex as the bait, see if we can recover the same proteins in the complex.
- Ho et al. Nature(2002) 415, 180-183;
  - 725 bait proteins in yeast; 1,578 interacting proteins FLAG tag, proteins transiently overexpressed
  - To assess confidence: 74% of interactions reproducible in small scale co-IP/blot

TAP/MS assay

- Pros
  - get the whole complex
  - proteins that purify together are likely to share a function
  - very sensitive -can detect ~15 copies per cell
  - *in vivo* conditions
  - can be adapted for high-throughput screens

TAP/MS assay

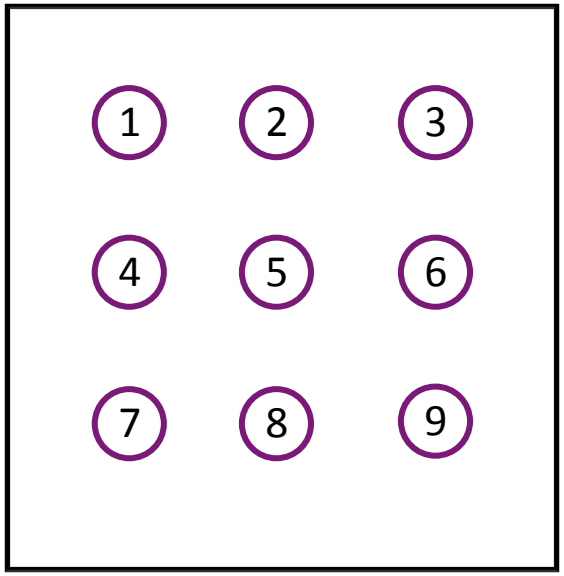
- Cons
  - doesn't determine direct or indirect interactions
  - not reliable for small proteins (< 15 kD)
  - affinity tag may interfere with interactions or with the function of essential proteins
  - prone to false positives, e.g. "sticky" proteins
  - prone to false negatives
    - won't get every protein every time
    - complex must survive purification
  - not quantitative

Experimental methods

- **Co-immunoprecipitation** is considered to be the gold standard assay for protein–protein interactions, especially when it is performed with endogenous (not overexpressed and not tagged) proteins.
- **Pull-down assays** are a common variation of immunoprecipitation and are used identically, although this approach is more amenable to an initial screen for interacting proteins.
- **Chemical cross-linking** is often used to "fix" protein interactions in place before trying to isolate/identify interacting proteins.
- **Yeast two-hybrid assay**
- **Tandem Affinity purification (TAP)**
- **Protein microarray**
- **Phage display**

Protein Microarray

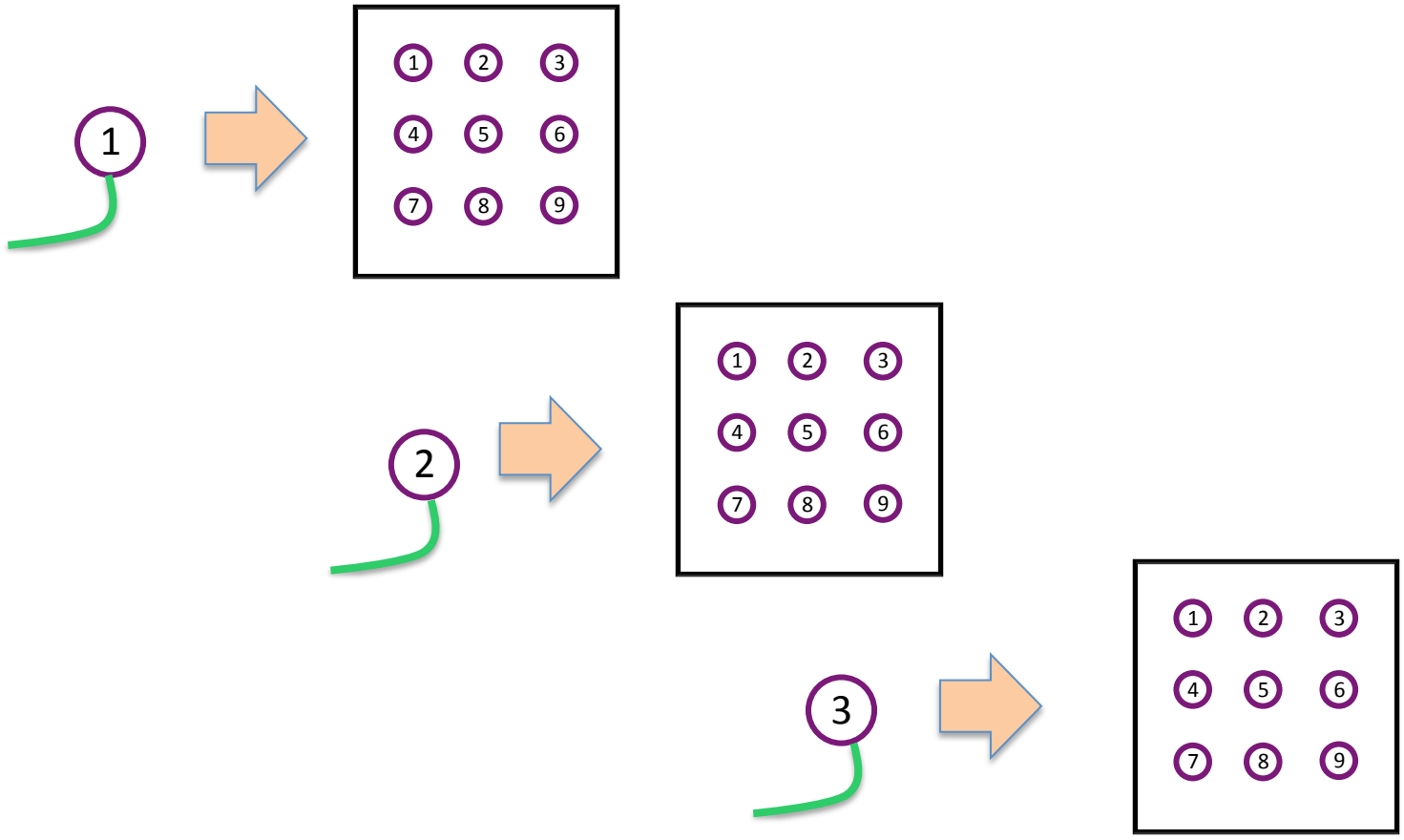
Making the whole genome protein on a microarray chip:



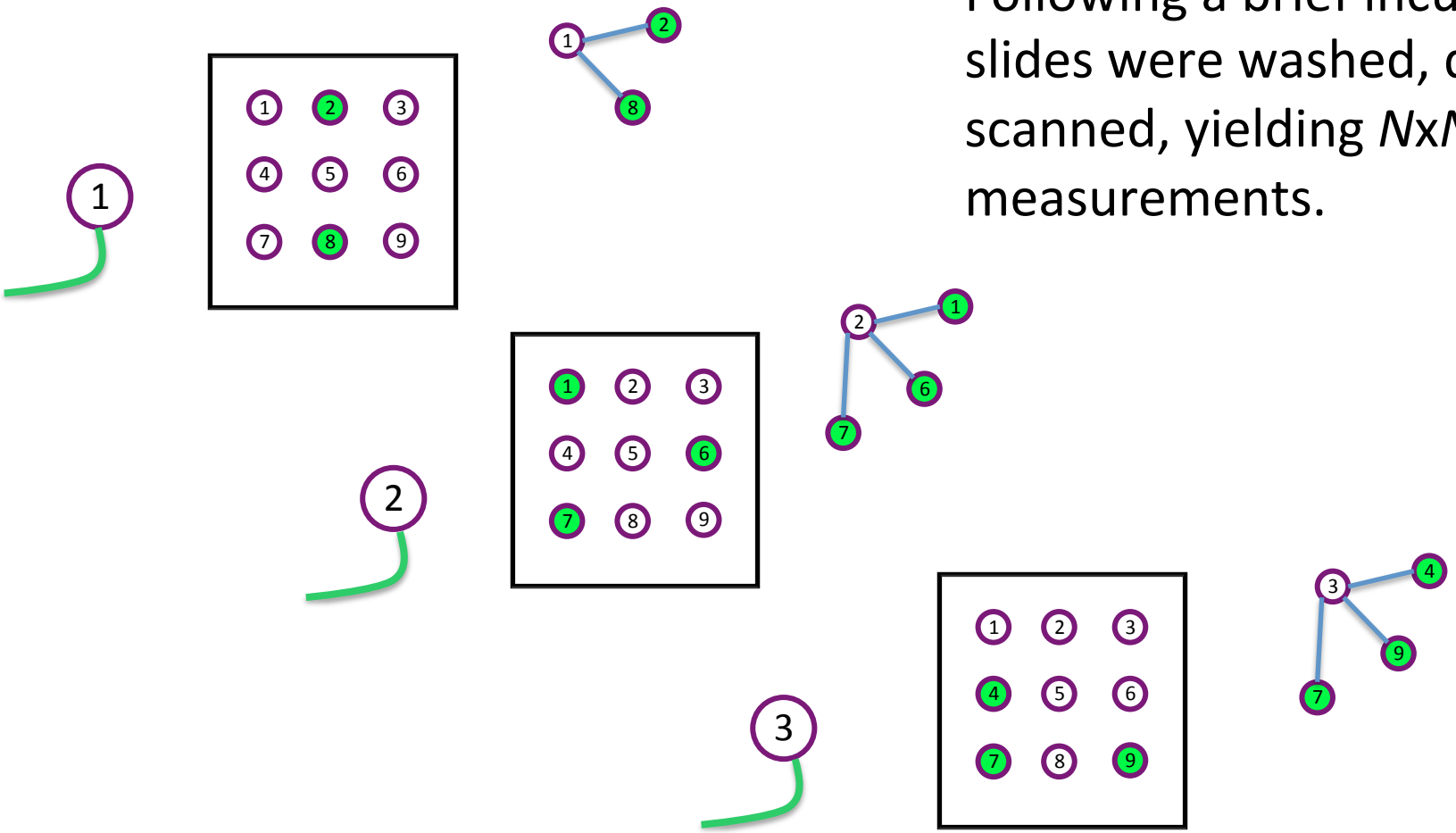
Highly purified proteins were denatured and printed onto glass slides.

Protein Microarray

- Labeling the purified protein with fluorescent dye, e.g. Cy-3.
- and add them to the arrays.

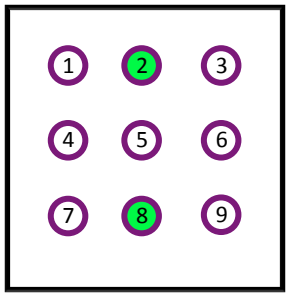


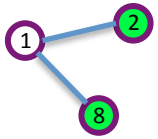
Protein Microarray

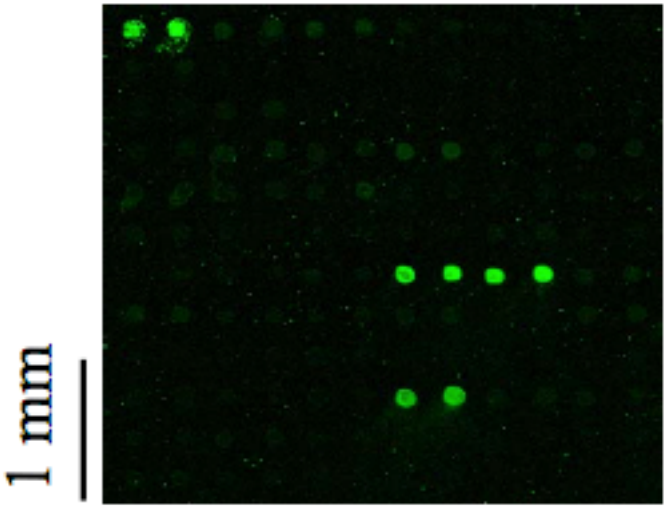


Following a brief incubation, slides were washed, dried and scanned, yielding *N*x*N* measurements.

Protein Microarray







Array Detection of Protein-Protein Interactions

- MacBeath & Schreiber Science 2000
  - proof-of-principle for three types of interactions protein-protein: protein G with IgG, FRAP with FKBP12, p50 with IκBα protein-small molecule: biotin with steptavidin, Ab with DIG steroid ligand enzyme-substrate: kinases PKA, Erk2
- Zhu *et al.* Science 2001
  - assay of 5,800 yeast genes with calmodulin, phospholipids
- Newman & Keating Science 2003
  - assay of ~48 x 48 human bZIP transcription factor coiled coils (plus 10 x 10 yeast)

Protein Microarrays

- Pros
  - Fast, N x N interactions at once
  - direct interaction assay
  - reagents can be well characterized
  - solution conditions are controlled
  - can be quantitative
  - requires very little protein
  - can be adapted for high-throughput screens
  - few false positives

Protein Microarrays

- Cons
  - tedious purification required, or else interactions may not be direct
  - surface may perturb folding or interactions
  - doesn't mimic *in vivo* conditions
  - not yet a mature technology -possibly not a good general approach, no commercial chip yet

Experimental methods

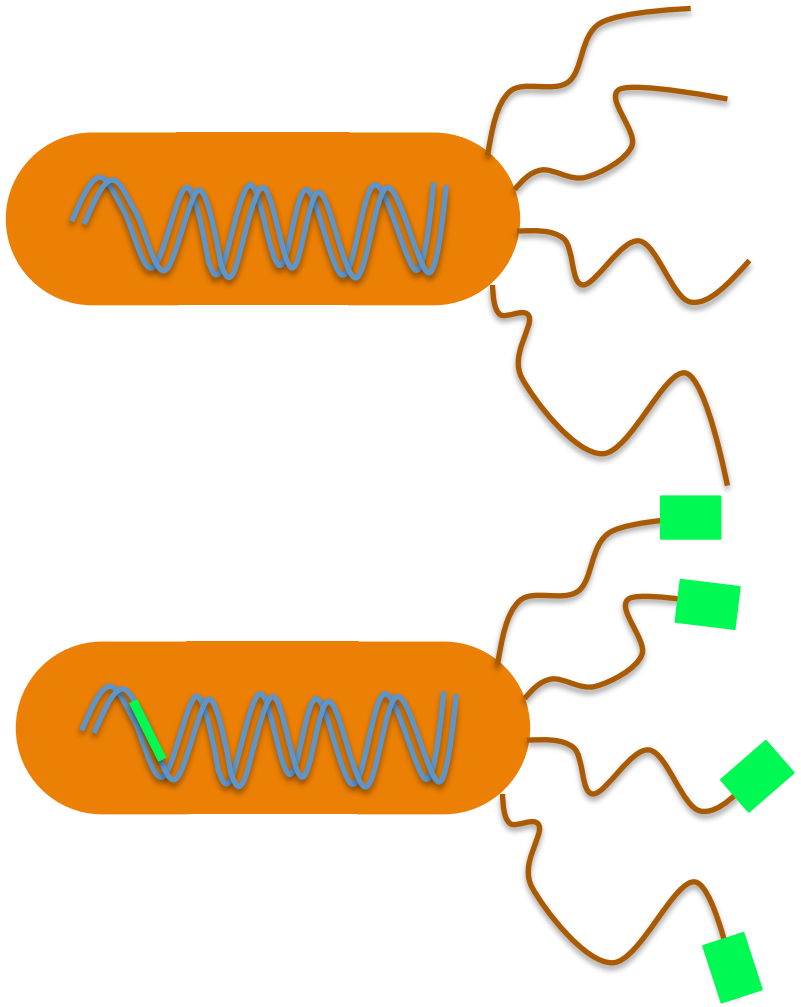
- **Co-immunoprecipitation** is considered to be the gold standard assay for protein–protein interactions, especially when it is performed with endogenous (not overexpressed and not tagged) proteins.
- **Pull-down assays** are a common variation of immunoprecipitation and are used identically, although this approach is more amenable to an initial screen for interacting proteins.
- **Chemical cross-linking** is often used to "fix" protein interactions in place before trying to isolate/identify interacting proteins.
- **Yeast two-hybrid assay**
- **Tandem Affinity purification (TAP)**
- **Protein microarray**
- **Phage display**

Phage display

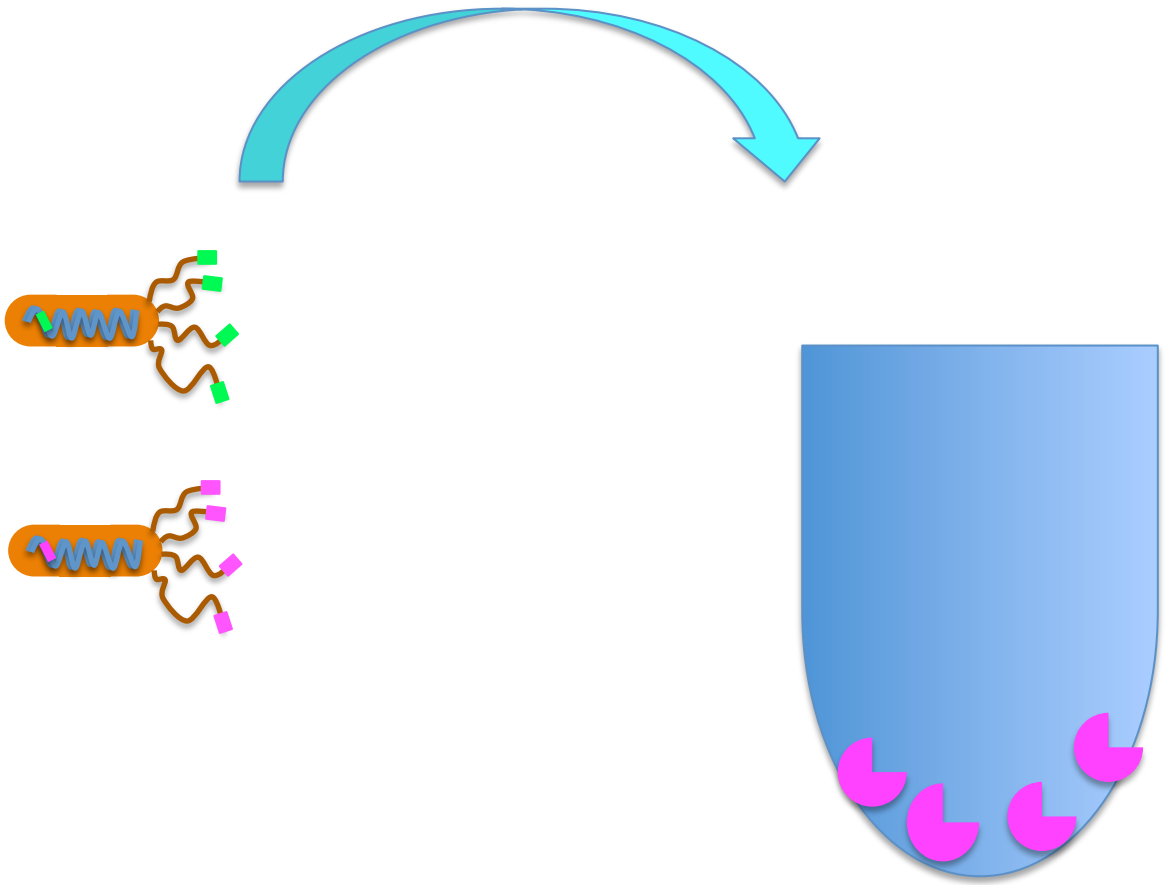
- **Phage display** is a method for the study of protein–protein and protein–DNA interactions that uses bacteriophages to connect proteins with the genetic information that encodes them.

Phage display

For M13 filamentous phage, the DNA encoding the protein or peptide of interest is ligated into the pIII or pVIII gene, encoding either the minor or major coat protein

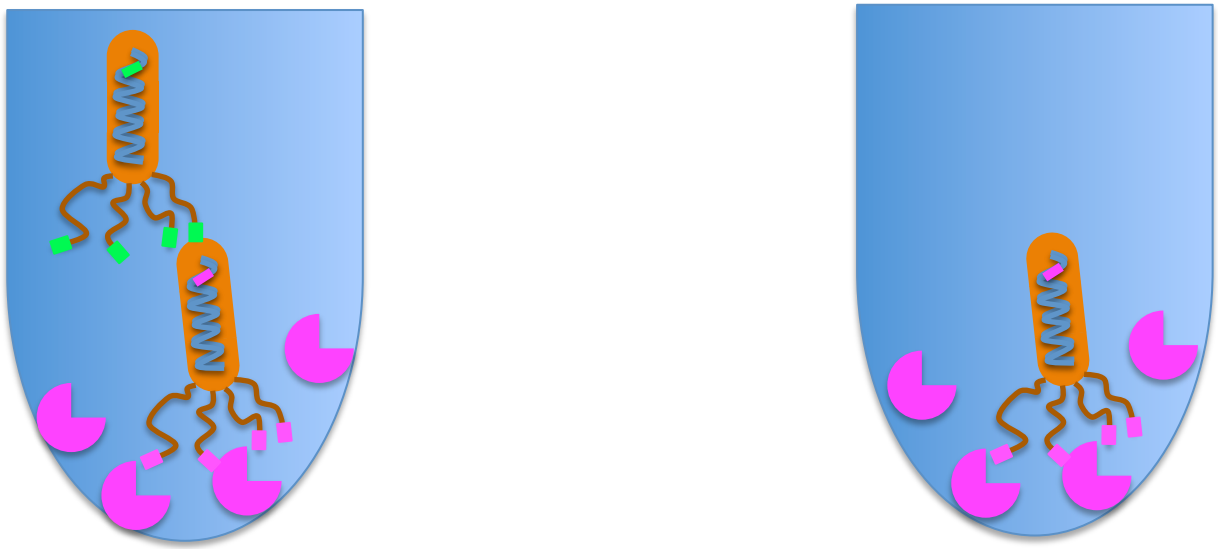


Phage display

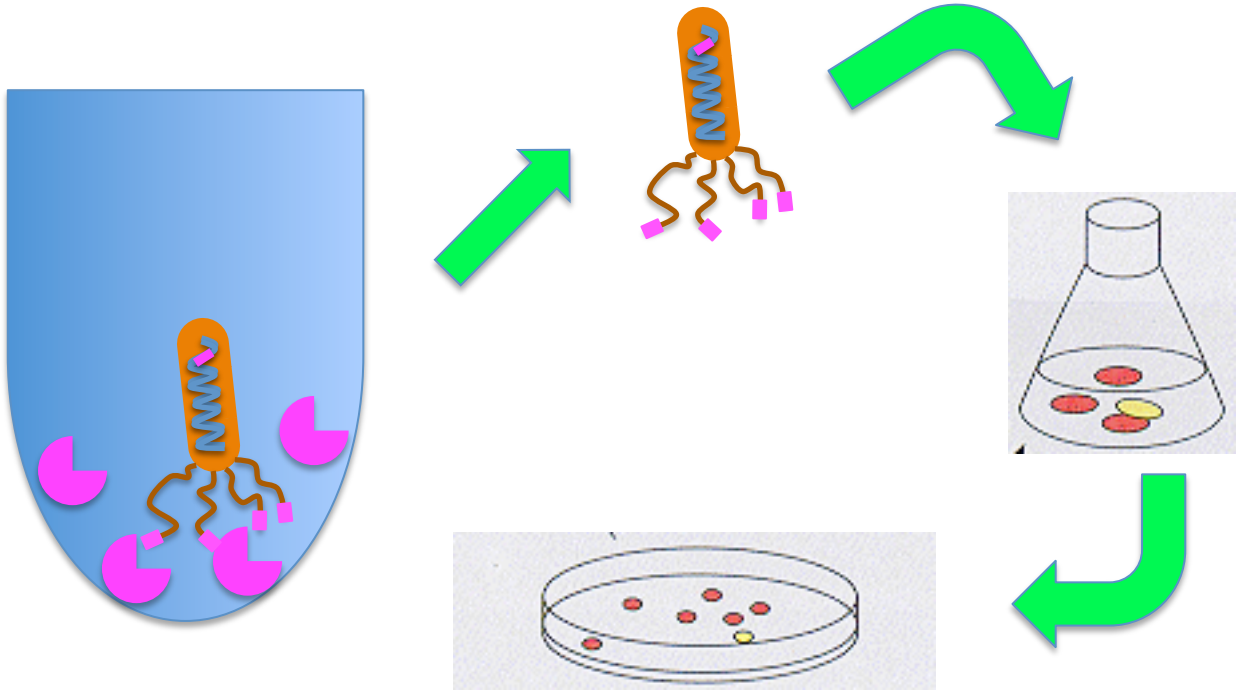


Immobilizing protein targets to the surface of a well

Phage display



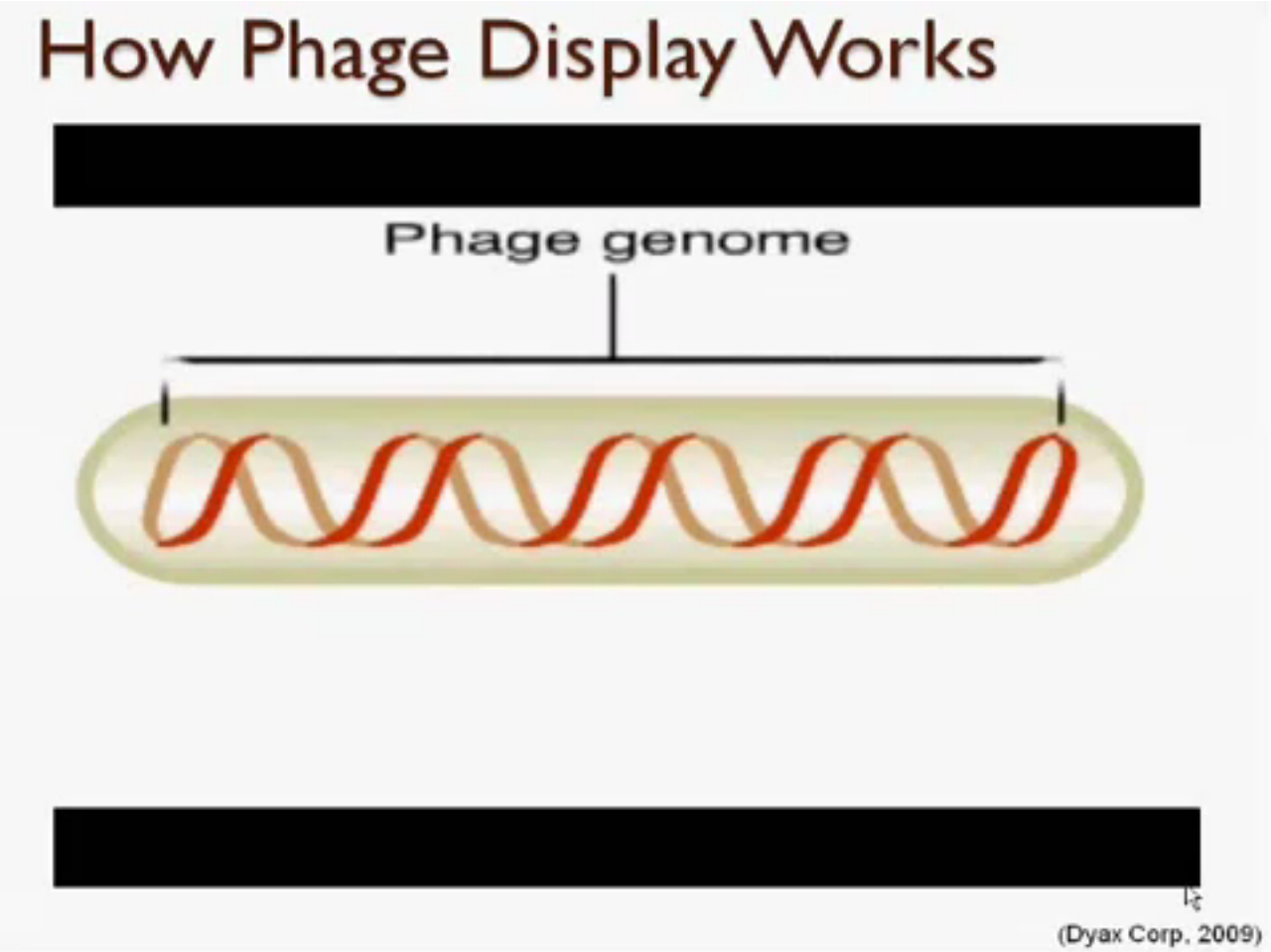
Phage display



Infect E. Coli

Phage eluted in the final step can be used to infect a suitable bacterial host, such as E. Coli, from which the “phagemids” can be collected and the relevant DNA sequence excised and sequenced to identify the relevant, interacting proteins or protein fragments.

Phage display video



http://www.youtube.com/watch?v=RuROQSlCz18

Phage Display

- Pros
  - No information of target protein needed
  - Can identify natural and non-natural ligands
  - Output of method is the DNA sequence of ligands
  - Fairly low false positive rate
  - can be adapted for high-throughput screens

Phage Display

Cons:

- Requires purified target protein
- Tightest binding phage may not represent biological partners
- Difficult to assay all sequence space
- Very *in vitro*

Overlap of high-throughput interaction studies is LOW

	Ito Y2H	Uetz Y2H	Gavin TAP/ms	Ho FLAG/ms
Ito 2-hybrid	4363	186	54	63
Uetz 2-hybrid		1403	54	56
Gavin affinity			3222	198
Ho affinity				3596
Small scale	442	415	528	391

data from Salwinski & Eisenberg, Current Opinion in Structural Biology (2003) 13, 377-382

Conclusions

- Lots of protein-protein interaction data are now available for yeast, but it is not very reliable and not comprehensive.
- Need additional accessing and filtering steps.
- Nevertheless, these data have inspired the development of many computational methods.
- To facilitate computational analysis, need to disseminate the data in a usable form! This is often a rate limiting step in systems biology.