Interaktionen von RNAs und Proteinen

Sonja Prohaska

Computational EvoDevo group
Universität Leipzig

SS2018

(Multi) Protein Complexes

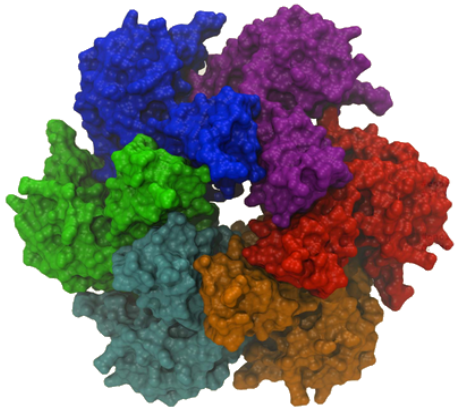
- two or more associated polypeptide chains
- **with homologous structure/function**

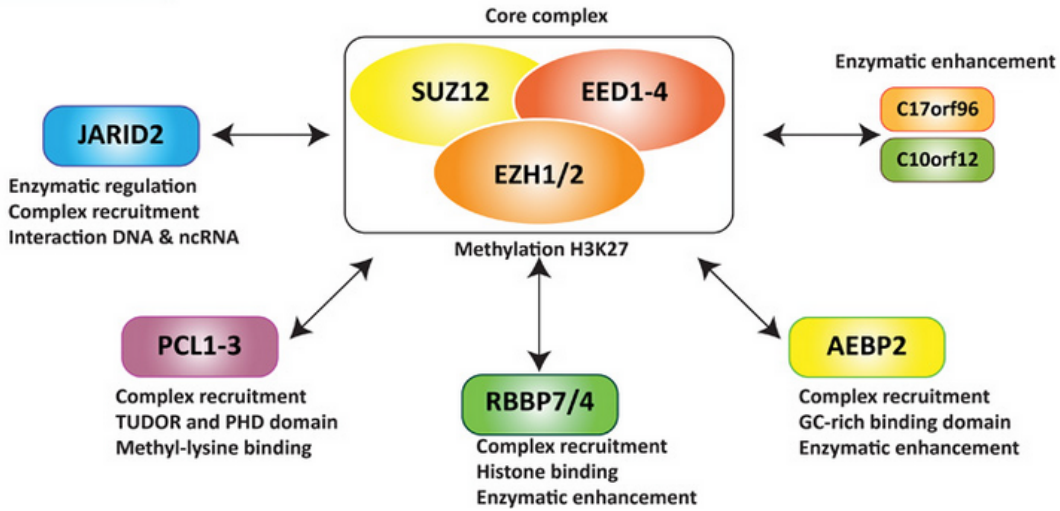
  formation of homo- and heterodimers to oligomers from monomers

  e.g. helicase
- **with different structure/function**

  built from (core) subunits and accessory proteins

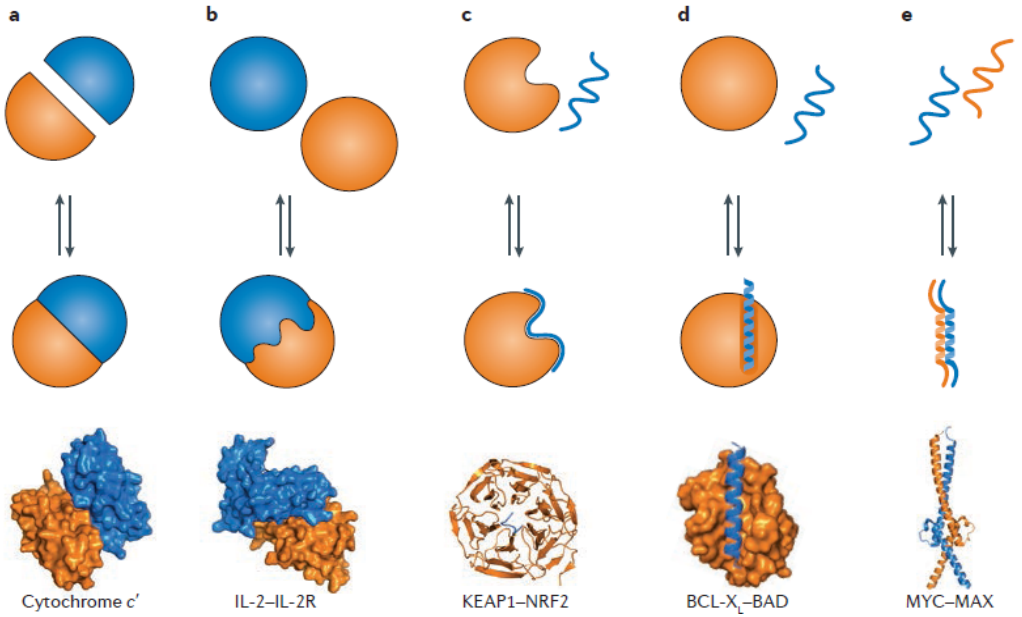
  e.g. PRC2 complex
- protein complexes are a form of quaternary structure
- strong bonding is based on disulphid bridges, salt bridges, hydrophobic contacts, electon sharing





Protein-Protein Interactions

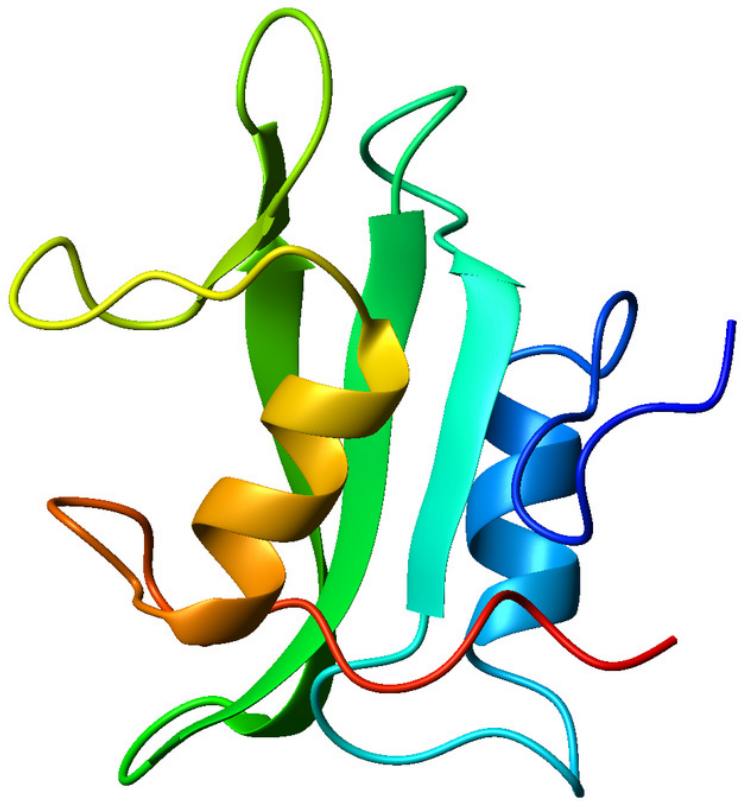
- define interaction interfaces/surfaces
- bonding is based on hydrophobic contacts, Van der Waals forces, hydrogen bonds



a two globular proteins with preformed surfaces
b two globular proteins with an induced binding surface
c rigid globular protein with a peptide
d flexible globular protein with a peptide
e interaction of two peptides

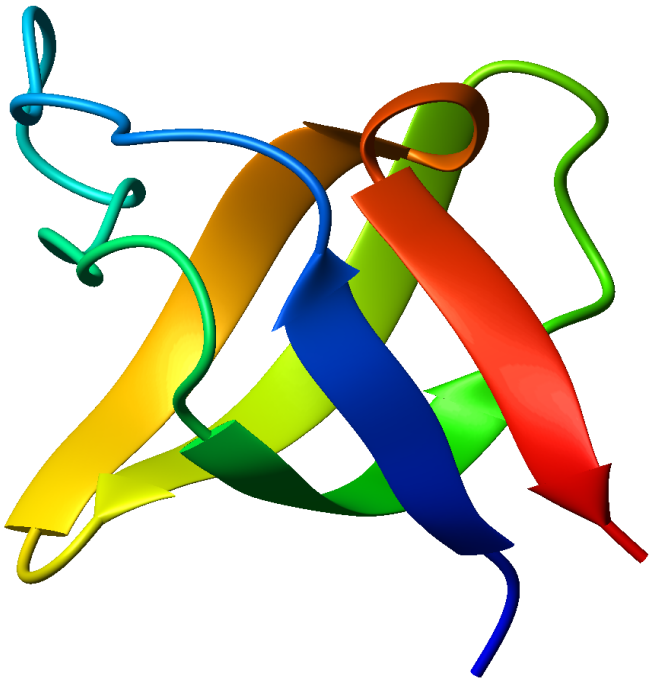
Protein-Protein Interaction Domain: SH2 Domain

- **S**rc **H**omology **2**
- around 100 aa
- 2 α-helices and 7 β-strands
  (i.e. one large β-sheet)
- known to identify a sequence of 3-6 aa
- high affinity to phosphorylated tyrosine
- function – signaling (transient binding)
- found in about 100 human proteins



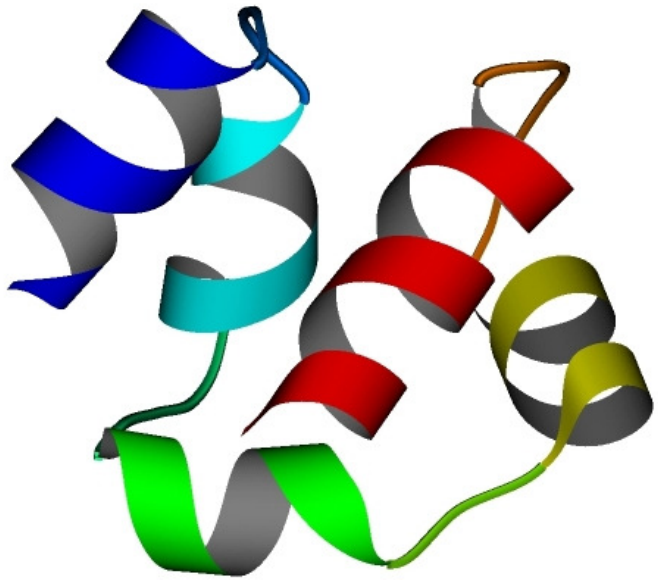
Protein-Protein Interaction Domain: SH3 Domain

- **S**rc **H**omology **3**
- about 60 aa
- beta-barrel fold: six β-strands forming two tightly packed anti-parallel β-sheets
- contacts proline-rich peptide sequence: -X-P-p-X-P- (X – aliphatic amino acid; p – sometimes proline)
- function – signaling (transient binding)
- found in about 300 human proteins



Protein-Protein Interaction Domain: SAM Domain

- **S**terile **A**lpha **M**otif
- around 70 aa
- small five-helix bundle
- has two large interfaces
- can form u.a. oligomers
- (some can bind RNA)
- found in about 60 human genes

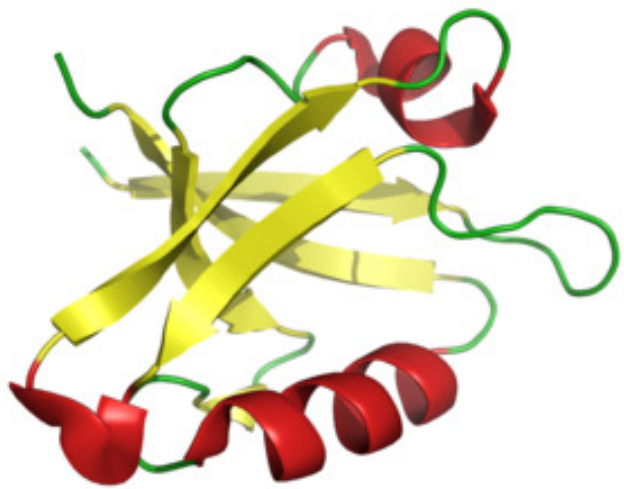


Protein-Protein Interaction Domain: PDZ domain

- 80-90 aa
- 5 β-sheets, some helices
- binds to C-terminus of binding partner

  by adding a β-strand to the β-sheet
- multiple PDZs per protein

  increase specificity
- function – localizing cellular elements and regulating cellular pathways
- 260 PDZ in 180 human genes



Tools

- InterPro: domain structure of proteins
  `https://www.ebi.ac.uk/interpro/`
- STRING: functional protein association networks
  `https://string-db.org/`

Yeast Two-Hybrid (Y2H)

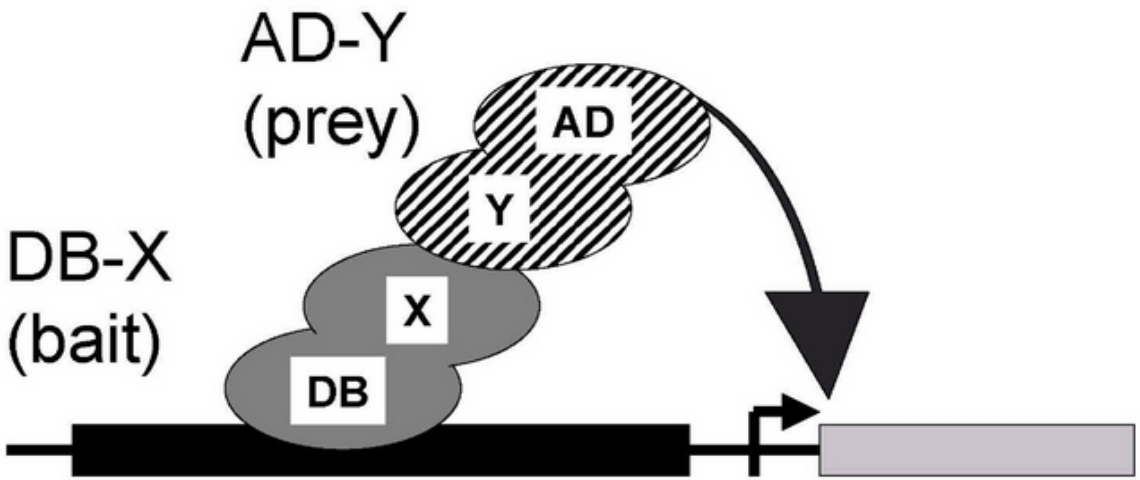
- test if X binds Y

  express fusion protein X-DB (DB ... DNA binding domain)

  express fusion protein Y-AD (AD ... activation domain)

  long linkers between X and DB, and Y and AD
- if X binds Y: reporter gene is expressed
- else: reporter gene is not expressed

  reporter gene: HIS3 (essential for histidine production)



Yeast *Saccharomyces cerevisiae*

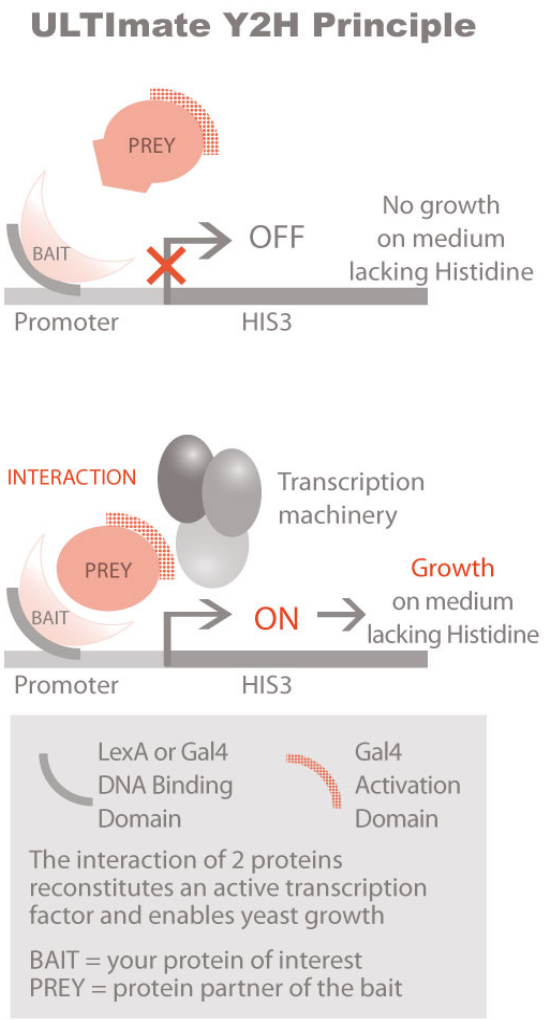
- genome size: 12.5×10^6bp
- about 5770 genes
- about 6100 proteins (about 2000 uncharacterized)
- 5100 soluble: 47% cytoplasm, 27% nucleus
- haploid and diploid living forms
- haploid types *a* and α can mate with each other

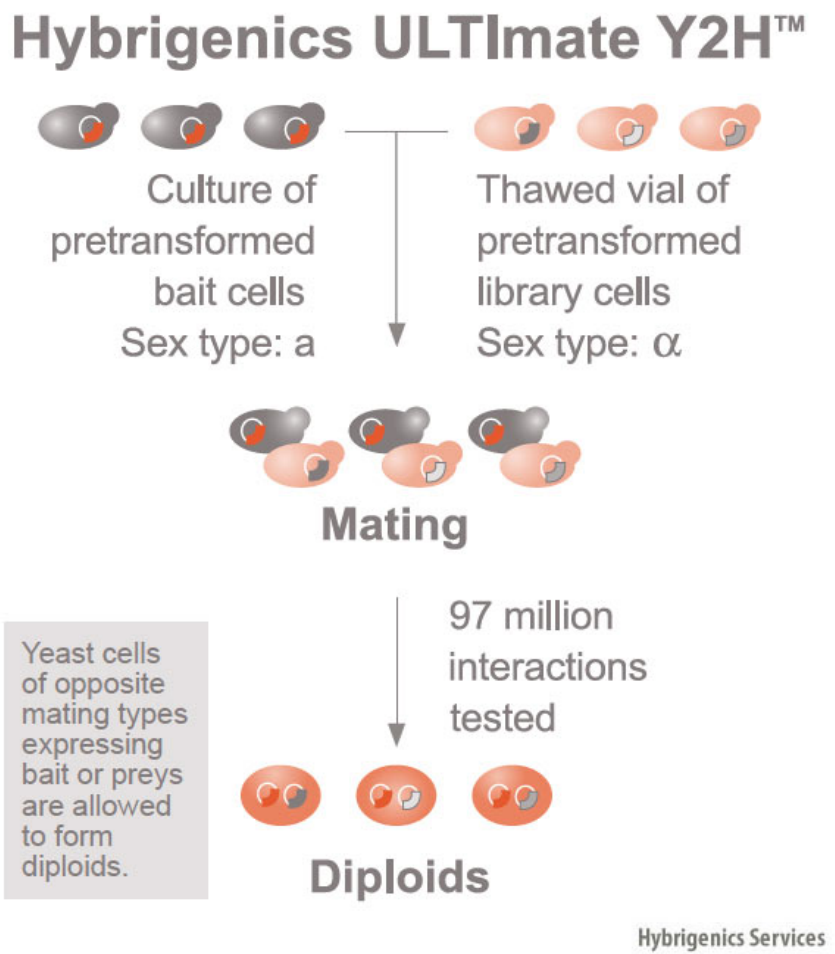
High-Throughput Yeast Two-Hybrid screening (HT-Y2H)

- a *bait* library (e.g. about 200 proteins)
- theoretically any number of bait
- a *prey* library (e.g. about 6000 proteins)
- *bait* encodes fusion protein: DB and protein X
- *prey* encodes fusion protein: AD and protein Y
- set of (imobilized) *bait*-expressing haploid cells (a)
- set of *prey*-expressing haploid cells (α)
- mate *bait* with *prey*
- retrieve diploid cells expressing reporter gene due to protein-protein interaction (PPI)

It is difficulty to sample **all** possible binary combinations of proteins using the library screening methods.

High-Throughput Yeast Two-Hybrid screening (HT-Y2H)





High-Throughput Yeast Two-Hybrid screening (HT-Y2H)

How well does it work?

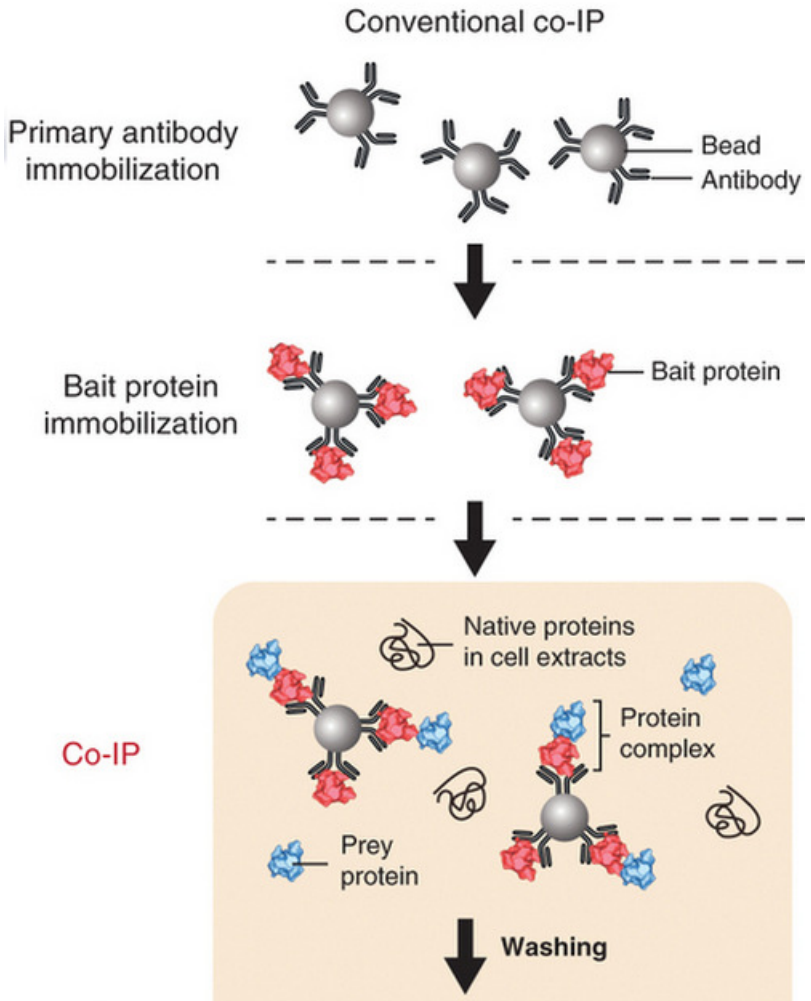
- 1-30 positives per *bait*
- only 20% true positives
- reason: self activation of reporter gene
- solution: two independent screens per bait
- even better solution: varify with other method
- found 87 baits in 281 PPI
- very low resolution
- Problem: construction of artificial fusion proteins

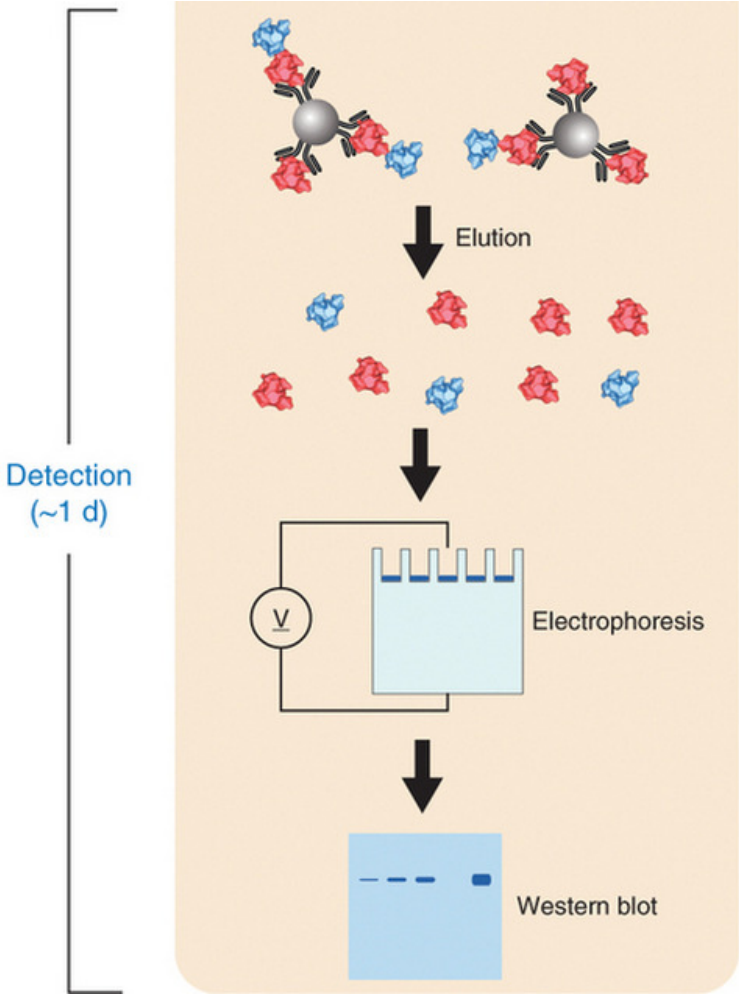
co-Immunoprecipitation (co-IP)

Co-Immunoprecipitation is not a high-throughput method.
It is used for validating PPI predictions, e.g. from HT-Y2H.

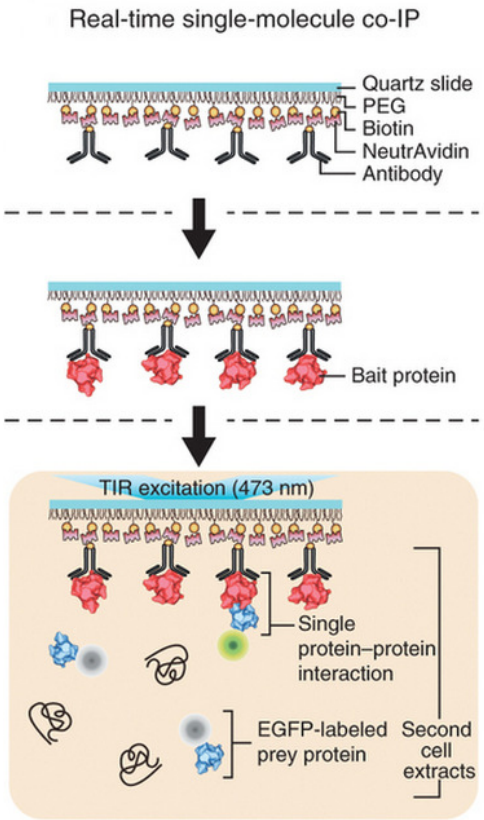
- antibodies against epitope on known protein (e.g. *bait*)
- pull the entire protein complex ("pull-down")
- works if proteins bind to each other tightly
- Problem: protein might hide epitope in complex
- Solution: antibody against different epitope on same protein
- Solution: antibody against proposed binding partner (double-check)
- Problem: no detection of transient interactions

co-Immunoprecipitation (co-IP)

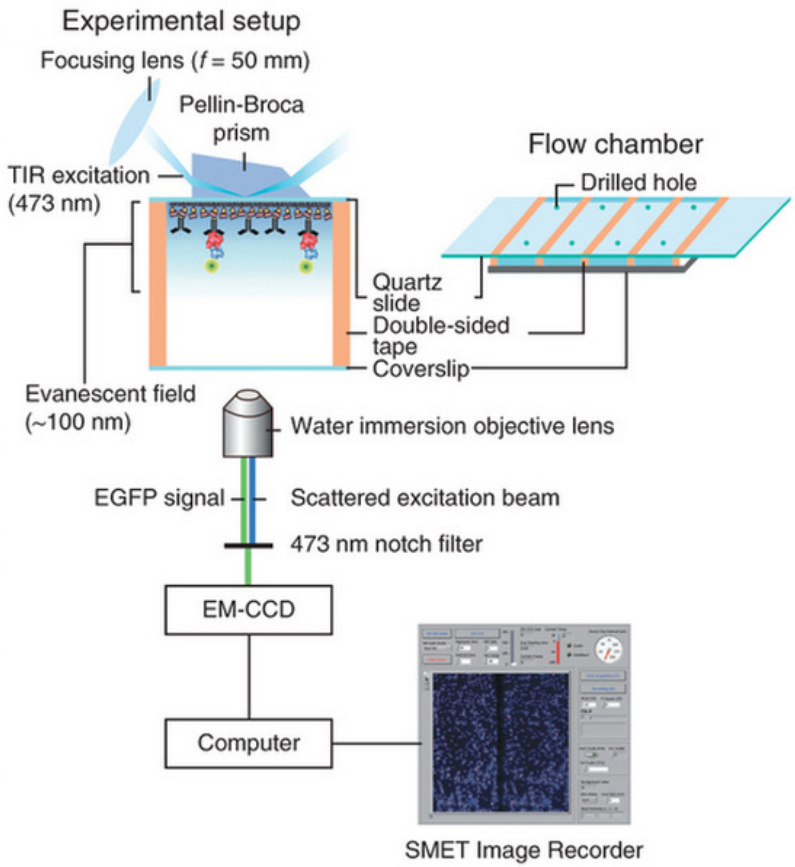




Real-Time Single-Molecule co-IP



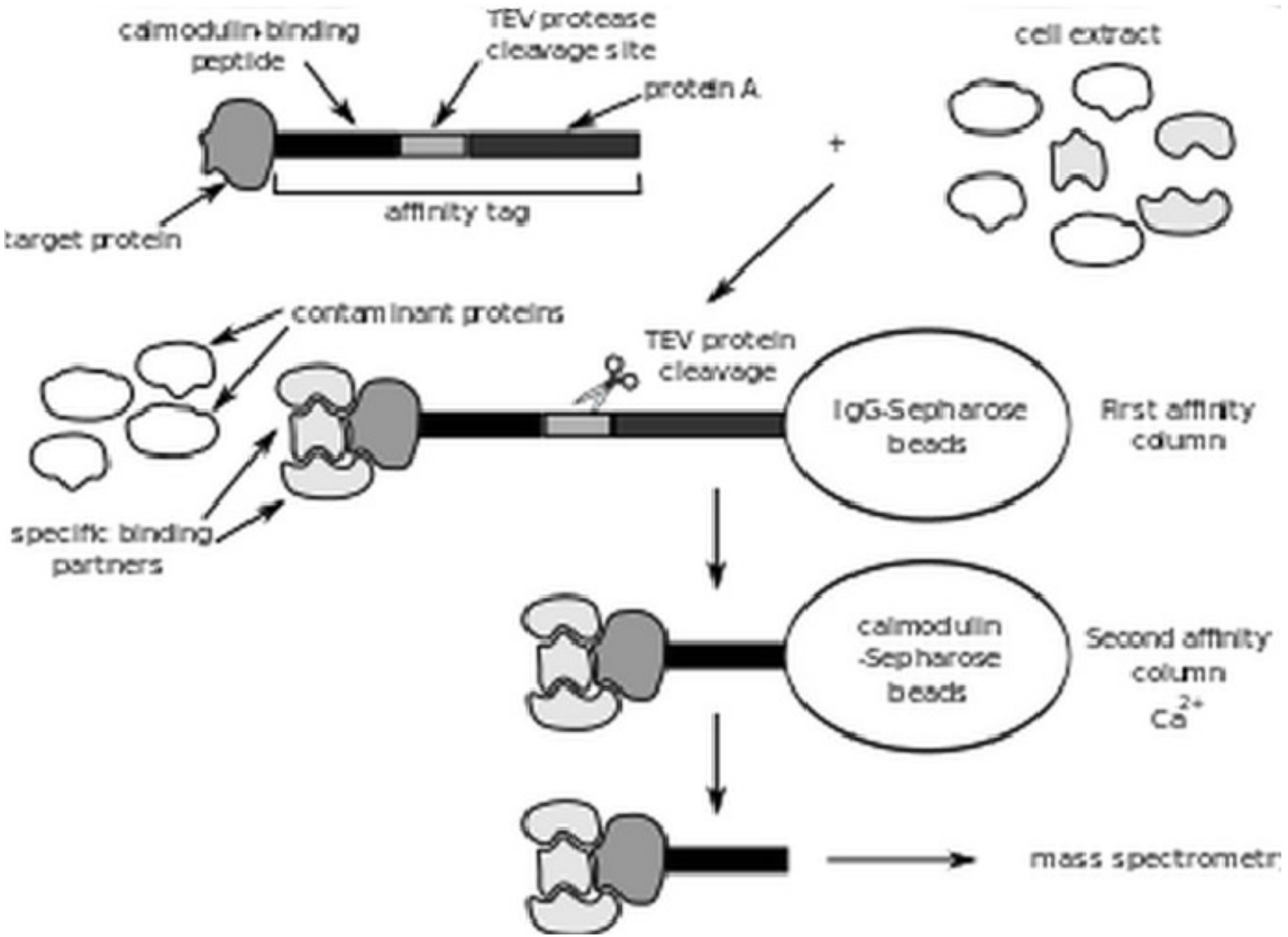
Co-IP
+
real-time
detection
(Time resolution
~50 ms)



Advantage: suited for transient interactions, kinetics studies (time resolution of 50ms)

Disadvantage: requires EGFP labeling of prey

Tandem Affinity Purification (TAP)



Tandem Affinity Purification (TAP)

- 'target' protein is tagged on C-terminus
- tag sequence: calmodulin binding peptide (CBP), cleavage site and protein A
- 1. purification: beads with IgG bind protein A, wash
- cleave at cleavage site, elute

  contamination is expectable $\rightarrow$
- 2. purification: beads with calmodulin bind CBP, wash
- analyse protein complexes with mass spectromentry

TAP and mass spectroscopy

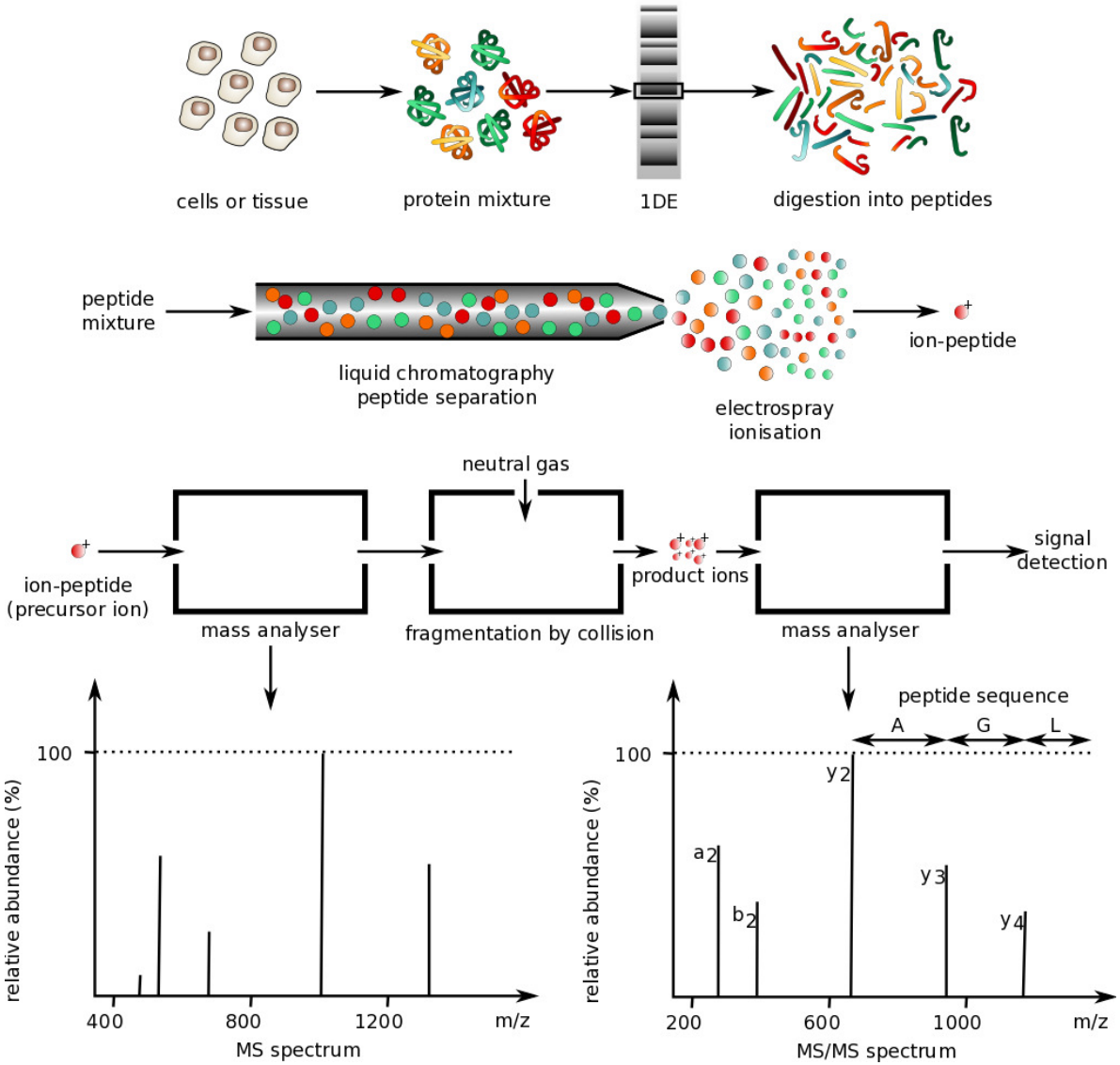
- using LC-MS, and MALDI/TOF-MS
- 4562 different tagged proteins
- 2708 proteins in 7123 PPI (2006)
- improvement!

Mass Spectrometry - General Idea

Identify and quantify multiple proteins in one run

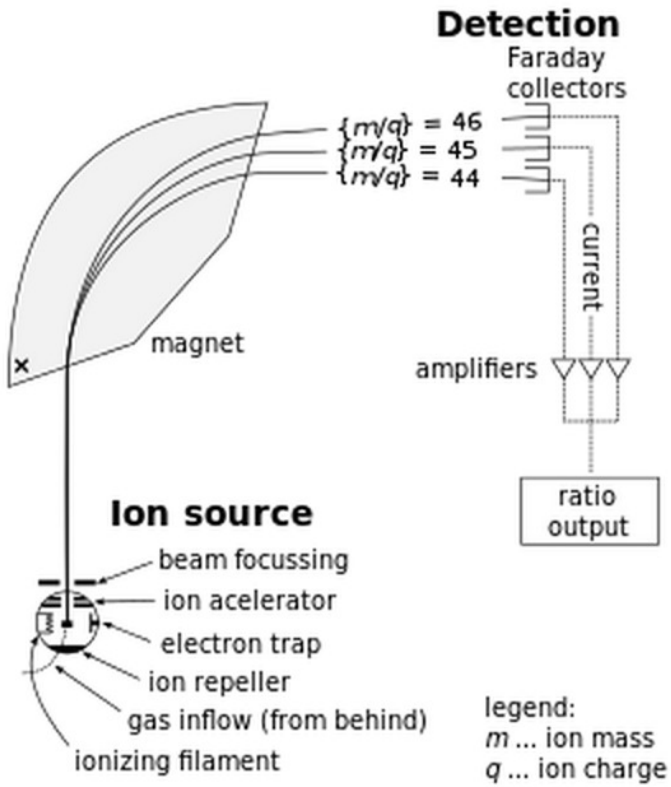
- starting with the pool of proteins
- coarse-grained protein separation by size (electrophoresis)
- protein fragmentation e.g. with trypsin (enzyme)
- fine-grained peptide separation with liquide chromatography (LC)
- electrospray ionization of peptides
- mass analysis $\rightarrow$ MS spectrum
- (collide peptides with neutral gas, even smaller fragments, mass analysis $\rightarrow$ MS/MS spectrum)
- **bioinformatic analysis** of MS spectrum

Protein Mass Spectrometry



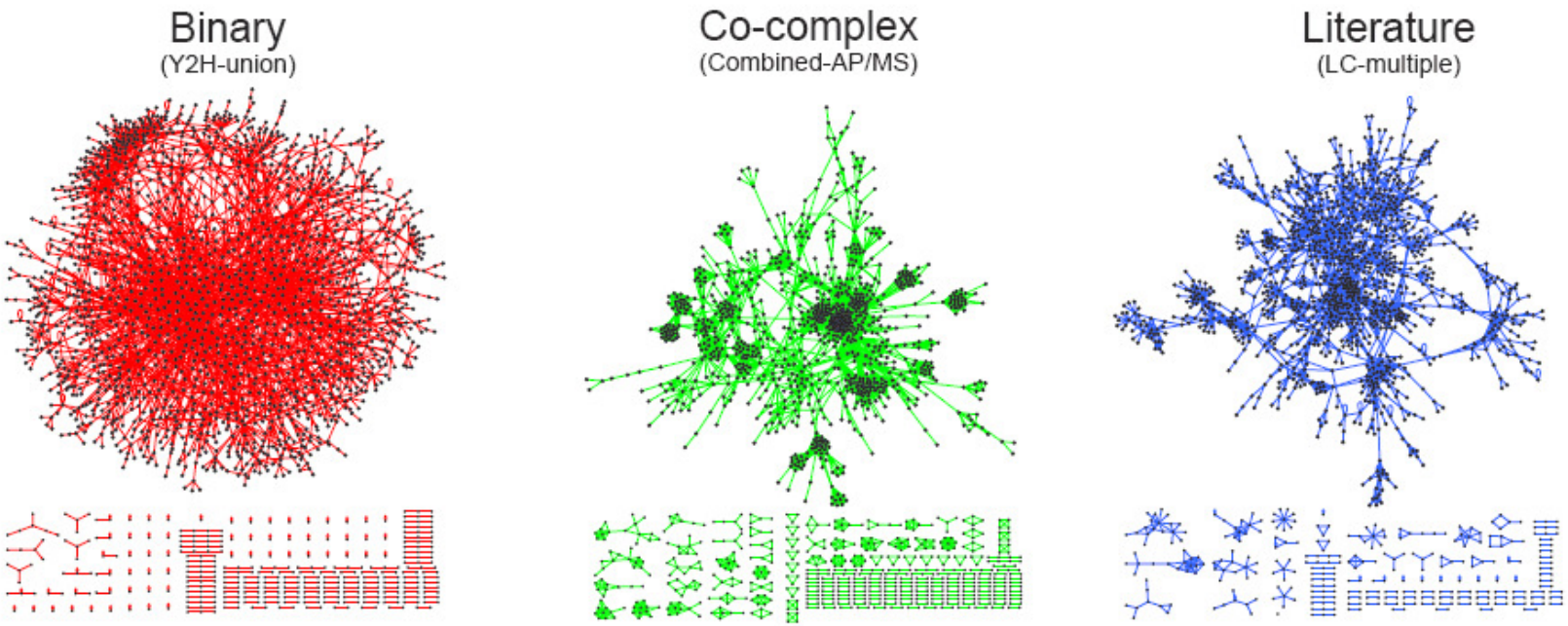
Mass Spectrometry

- mass (m) charge (q) ratio
- **peptide mass fingerprints** are stored in databases
- for comparison



- some amino acids have identical masses
- utilize a sequence homology search in parallel
- problem: modifications change mass $\rightarrow$ missidentification

Interactome



Propose “highly significant clustering between essential proteins”

Interactome as a Graph?

- only binary protein interactions are concidered
- intersting interactions might be transient
- only a static picture lacking dynamics and context
- A interacts with B and B interacts with C
  - at the same time?
  - in the same compartment or cell type?
  - Is there direct or indirect interaction of A and C?
  - Do A, B and C form a complex?
- would we see pathways?
- what can we infer from such a network?

Literature

Hong-Won Lee, Ji Young Ryu, Janghyun Yoo, Byungsan Choi, Kipom Kim, Tae-Young Yoon. *Real-time single-molecule coimmunoprecipitation of weak protein-protein interactions.* Nature Protocols 2013; 8, 2045-2060

For further reading:

`https://www.intechopen.com/books/protein-protein-interactions-computational-and-experimental-tools`