Interaktionen von RNAs und Proteinen

Sonja Prohaska

Computational EvoDevo
Universitaet Leipzig

SS17
(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, electron 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

- **Src Homology 2**
- around 100 aa
- 2 $\alpha$-helices and 7 $\beta$-strands
- (one large $\beta$-sheet)
- known to identify a sequence of 3-6 aa
- high affinity to phosphorylated tyrosine
- function – signaling
- found in about 100 human proteins
Protein-Protein Interaction Domain: SH3 Domain

- **Src Homology 3**
- beta-barrel fold: six $\beta$-strands forming two tightly packed anti-parallel $\beta$-sheets
- contacts proline-rich peptide sequence: $-X$-$p$-$X$-$P$- ($X$ – aliphatic amino acid; $p$ – sometimes proline)
- function – signaling
- found in about 300 human proteins
Protein-Protein Interaction Domain: SAM Domain

- Sterile Alpha Motif
- around 70 aa
- small five-helix bundle
- seems to possess the ability to bind RNA
- has two large interfaces
- can form dimers
- found in small group of genes
Protein-Protein Interaction Domain: PDZ domain

- 80-90 aa
- 5 $\beta$-sheets, some helices
- binds to C-terminus of binding partner by adding a $\beta$-strand to the $\beta$-sheet
- multiple PDZs per protein increase specificity
- 260 PDZ in 180 human genes
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
Yeast *Saccharomyces cerevisiae*

- genome size: $12.5 \times 10^6$bp
- about 5770 genes
- about 6100 proteins (about 2000 uncharacterized)
- 5100 soluble: 47% cytoplasm, 27% nucleus
- haploid and diploid living forms
High-Throughput Yeast Two-Hybrid screening (HT-Y2H)

- a *bait* library (e.g. 192 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
- array of *prey*-expressing haploid cells
- library of *bait*-expressing haploid cells
- 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)

How it works: see blackboard

- 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)

Conventional co-IP:
- Primary antibody immobilization
- Bait protein immobilization
- Native proteins in cell extracts
- Protein complex
- Prey protein
- Washing

Detection (~1 d):
- Elution
- Electrophoresis
- Western blot
Real-Time Single-Molecule co-IP

**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: calmodulin binding peptide (CBP), cleavage site, protein A
- 1. purification: beads with IgG bind protein A, wash
- cleave at cleavage site, elute
- 2. purification: beads with calmodulin bind CBP, wash
- analyse protein complexes with mass spectrometry

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
- fine-grained peptide separation with lique chromatography (LC)
- electrospray ionization of peptides
- mass analysis → MS spectrum
- (collide peptides with neutral gas, even smaller fragments, mass analysis → MS/MS spectrum)
- **bioinformatic analysis** of MS spectrum
Protein Mass Spectrometry

1. Cells or tissue
2. Protein mixture
3. Digestion into peptides
4. Peptide mixture
5. Liquid chromatography peptide separation
6. Electrospray ionisation
7. Ion-peptide
8. Mass analyser
9. Fragmentation by collision
10. Product ions
11. Mass analyser
12. Signal detection

Relative abundance (%) vs. m/z

MS spectrum

MS/MS spectrum

peptide sequence

A G L
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 → misidentification
Propose “highly significant clustering between essential proteins”
Interactome as a Graph?

- only binary protein interactions are considered
- interesting 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?

For further reading: https://www.intechopen.com/books/protein-protein-interactions-computational-and-experimental-tools