Interaktionen von RNAs und Proteinen

Sonja Prohaska

Computational EvoDevo
Universitaet Leipzig

June 2, 2017
Stable (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: disulphid bridges, salt bridges, hydrophobic contacts, electron sharing
(Transient) Protein-Protein Interactions

- define interaction interfaces/surfaces
- loose bonding: hydrophobic contacts, Van der Waals forces, hydrogen bonds
- five structural classes of interactions
  - 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**
- 2 α-helices and 7 β-strands
- 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-p-X-P- (X being an aliphatic amino acid)
- 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: verify 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)

1. Primary antibody immobilization
2. Bait protein immobilization
3. Conventional co-IP
4. Native proteins in cell extracts
5. Protein complex
6. Prey protein
7. Washing
8. Elution
9. Detection (~1 d)
10. Electrophoresis
11. 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 liquid 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

1. Cells or tissue
2. Protein mixture
3. 1DE (Isolation of Proteins)
4. Digestion into peptides
5. Peptide mixture
6. Liquid chromatography (Peptide separation)
7. Electrospray ionisation
8. Ion-peptide
9. Neutral gas
10. Mass analyser
11. Fragmentation by collision
12. Product ions
13. Mass analyser
14. Signal detection
15. Relative abundance (%) vs. m/z

MS spectrum:
- a2
- b2
- y2
- y3
- y4

MS/MS spectrum:
- 200
- 600
- 1000
- 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 → missidentification
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?
Literature


**For further reading:**