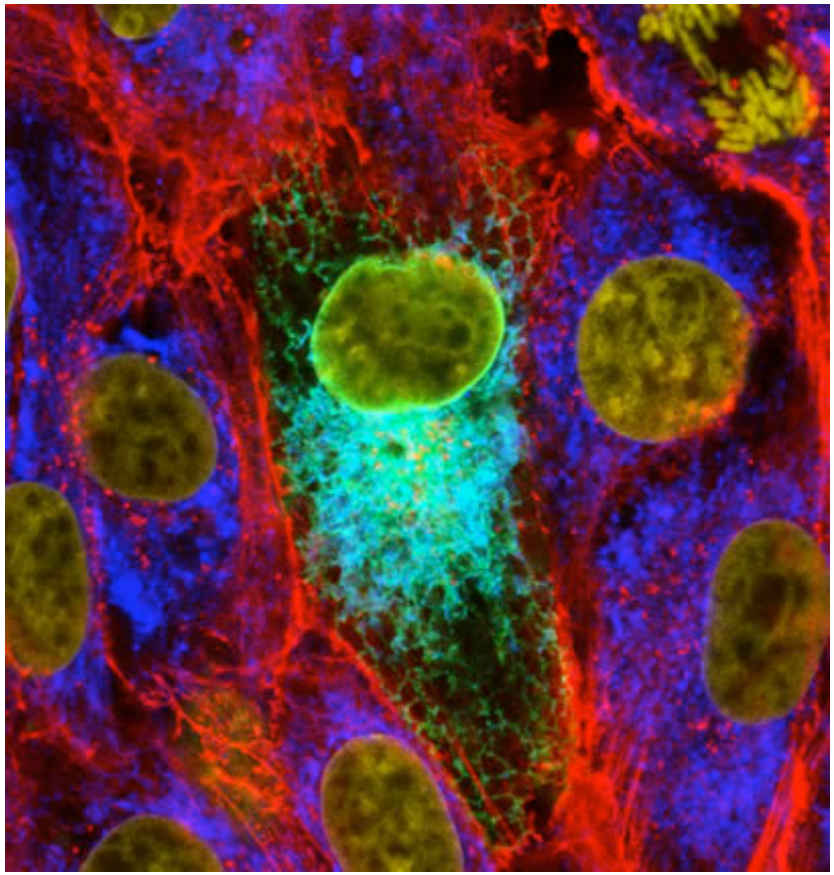
Protein Interactions: Complexes and Networks

Characterisation of Protein Interactions
Experimental Methods
Computational Methods
Interaction Networks

Crowded Cell



In the highly crowded environment of a cell, biological macromolecules occur at a concentration of 300-400 g/l and they physically occupy a significant fraction of the total volume (20-30%).

Most proteins interact at least transiently with other proteins, but it is clear that many essential cellular processes such as signal transduction, transport, cellular motion and regulatory mechanisms are mediated by protein-protein interactions.

Types of protein-protein interactions (PPI)

Obligate PPI

the protomers are **not** found as stable structures on their own *in vivo*

Obligate homodimer

P22 Arc repressor DNA-binding

Obligate heterodimer

human cathepsin D

1LYB

Non-obligate PPI

Non-obligate homodimer

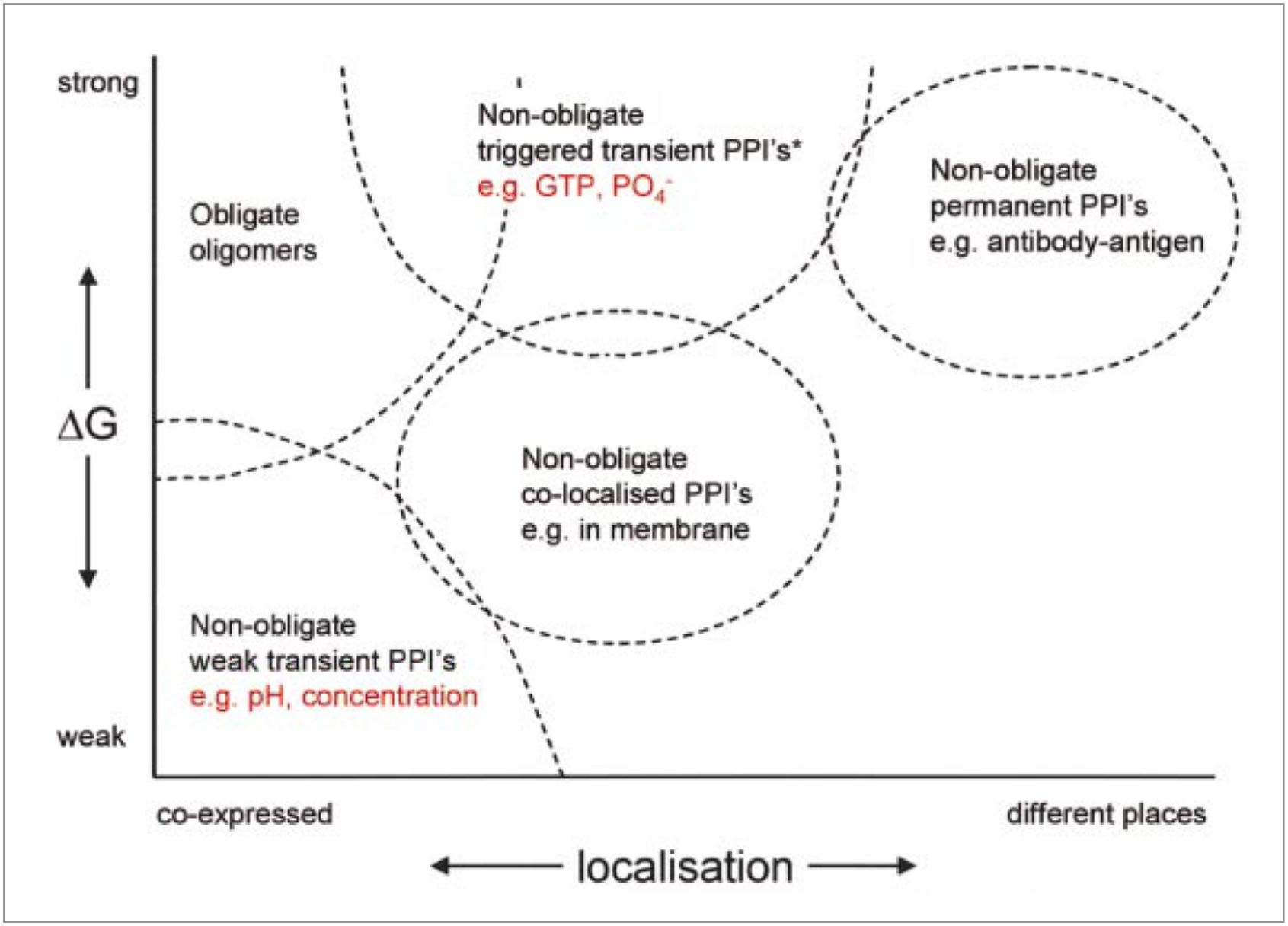
lysin

Non-obligate heterodimer

RhoA and RhoGAP signalling complex

Complex Classification

Complex Classification



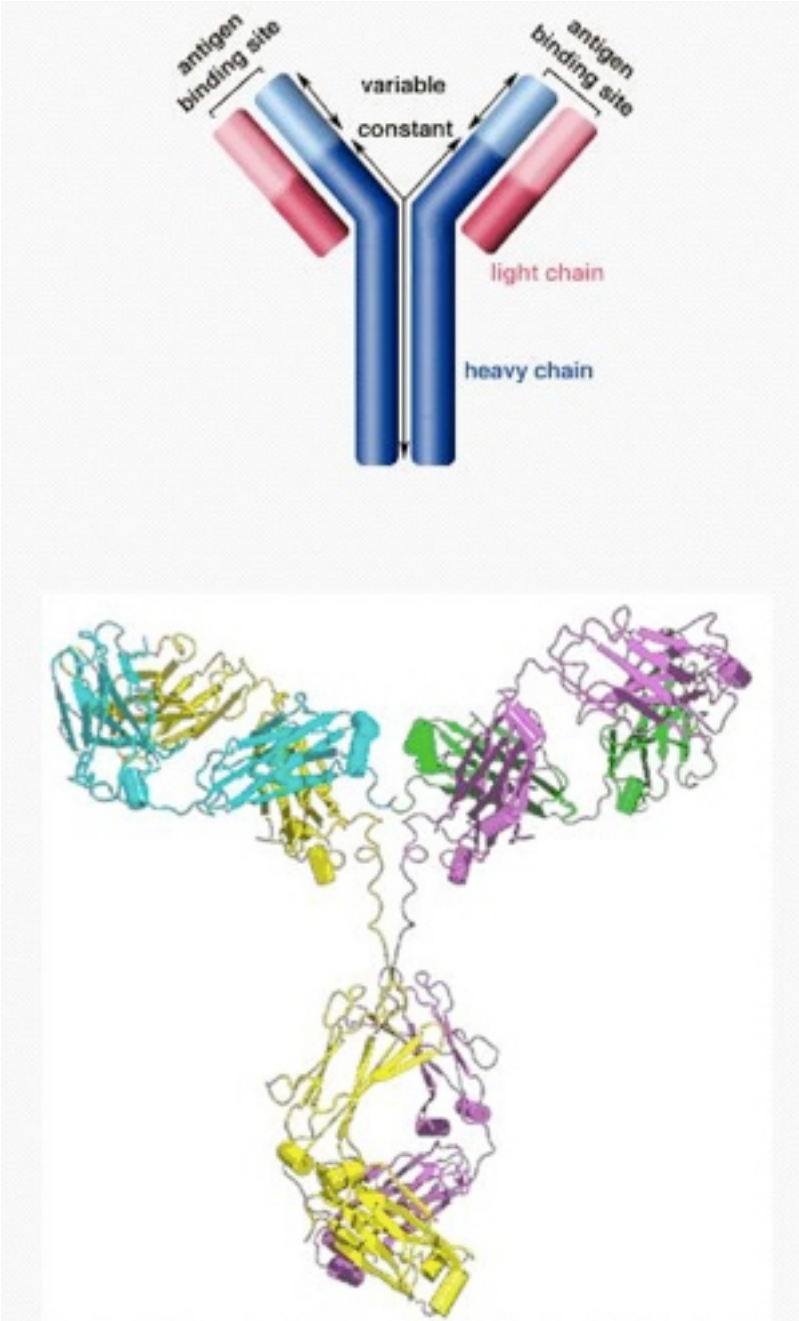
Interaction Types

1. Protein-Protein Interaction
   1.1. enzyme ⇔ substrate : serine protease trypsin
   1.2. receptor ⇔ ligand : growth hormone receptor
   1.3. antibody ⇔ antigen : immunoglobulin (Ig)

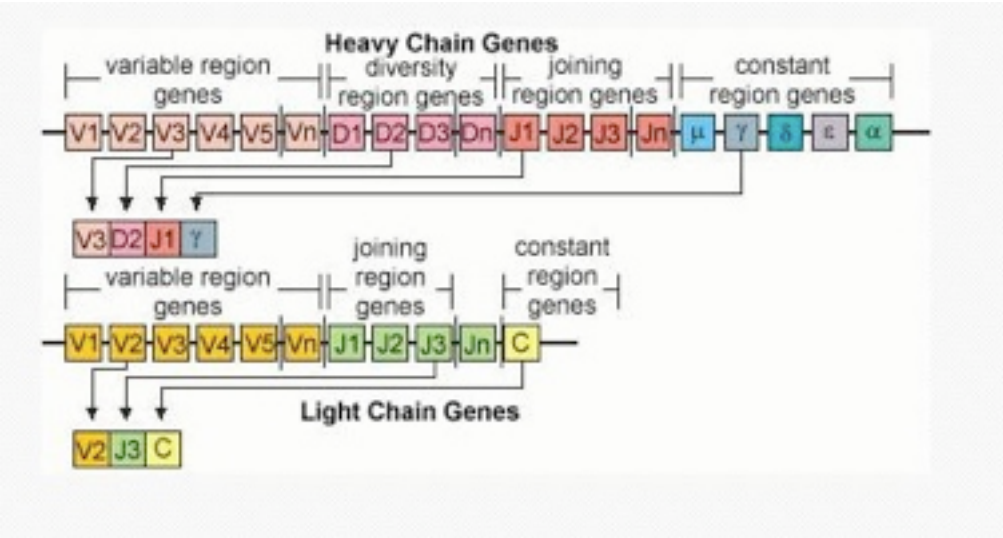
2. Protein-Ion and Protein-Small_Molecule Interaction
   2.1. protein ⇔ ion (Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Cl^{-}, HCO_3^{-}, SO_4^{2-}) : calmodulin
   2.2. pump ⇔ ion , coupled to enzymatic function : ATPase
   2.3. channel ⇔ water : aquaporin

3. Protein-DNA/RNA Interaction
   3.1. enzyme ⇔ DNA : Eco-RI ribozyme
   3.2. binder ⇔ DNA groove : leucine zipper, zinc finger
   3.3. regulator ⇔ RNA : KH domain

Antibody Structure



Antibody Diversity



Combination of genes is basis for diversity.

Antibodies are highly specific and show high affinity to antigen.

Antibodies bind with a combination of loops.

Complexation Reaction

Reaction Equation

$$A + B <=> AB \qquad K_a = \frac{c(AB)}{c(A)\,c(B)} \qquad \textit{association reaction}$$

$$AB <=> A + B \qquad K_d = \frac{c(A)\,c(B)}{c(AB)} \qquad \textit{dissociation reaction}$$

$$K_a = \frac{1}{K_d}$$

Free reaction enthalpy

$$\Delta G = -\,RT\,lnK$$

$$with\ R = k_B\,N_A = 8.3144\,Jmol^{-1}K^{-1}$$

1. Kinetics of intermolecular interactions

- electrostatic steering

Macromolecular association/dissociation constants:

Protein	Ligand	k_1	k_{-1}
Insulin	Insulin	1.2×10^8	1.5×10^4
Barnase	Barstar	$> 5 \times 10^9$	8.0×10^{-6}
Phenylalanyl-tRNA synthetase	tRNA	1.6×10^8	27

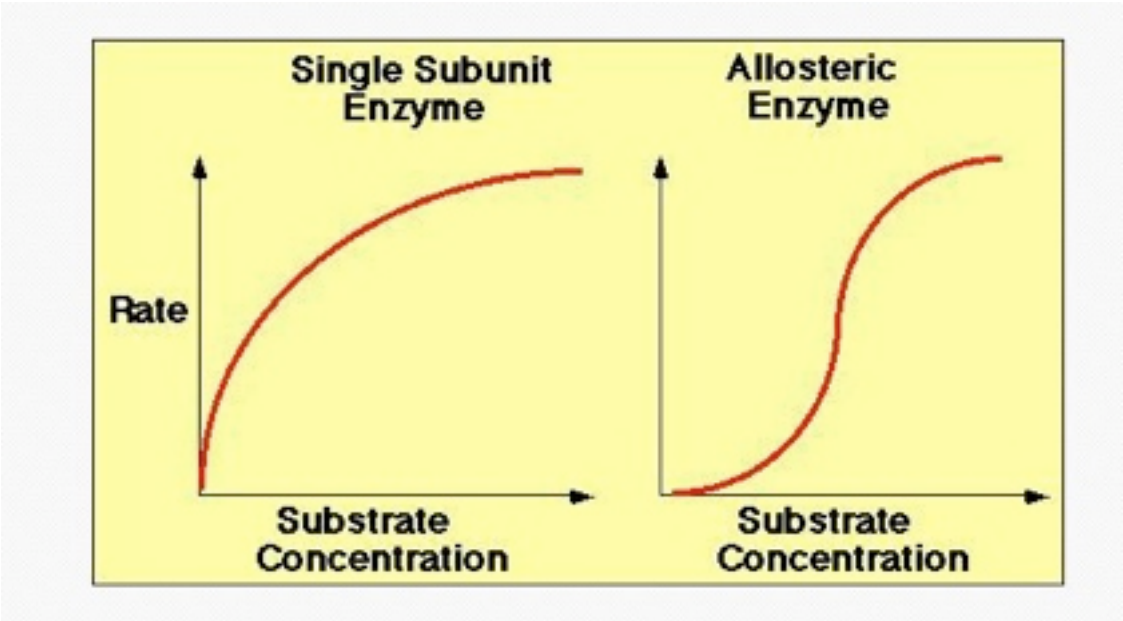
Enzyme/substrate association/dissociation constants:

Protein	Ligand	k_1	k_{-1}
Lactase dehydrogenase	NADH	Around 10^9	10^4
Catalase	H_2O_2	1.5×10^6	1
Chymptrypsin	Proflavin	1.2×10^8	8.3×10^3

Interface Area

- Contact area : usually >1100 Å^2, each partner >550 Å^2
- Solvent contact : On average each partner loses ~800 Å^2 of solvent accessible surface area, about 20 amino acids lose 40 Å^2
- Interaction energy : ~100-200 J per Å^2
- Buried accessible surface area (average):
  12% for dimers
  17% for trimers
  21% for tetramers
- Curvature : 83-84% of all interfaces are flat
- Secondary Structure : 50% α-helix, 20% β-sheet, 20% coil, 10% mixed
- Hydropathy : Less hydrophobic than core, more hydrophobic than exterior

Allostery

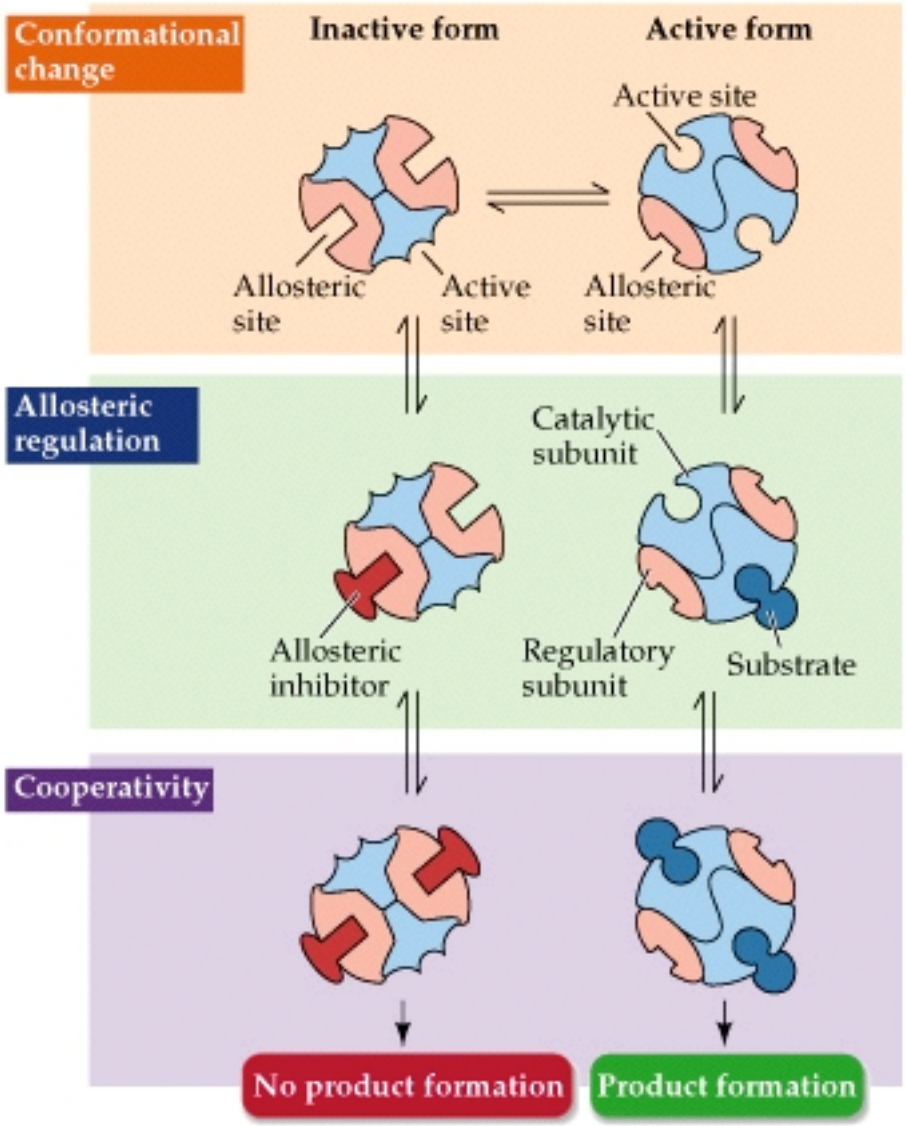


Binding of substrates or ligands is cooperative.

Domain changes conformation upon ligand binding; often in homodimers.

Cooperativity creates switch-like response.

Allostery



© 2001 Sinauer Associates, Inc.

Protein Complex Detection

Structure Analysis

Xray or NMR, fluorescence spectroscopy

Separation Techniques

2D gel electrophoresis and mass spectrometry

Chromatography (gel permeation)

Immobilisation / Binding Techniques

Surface Plasmon Resonance

Affinity Chromatography

Equilibrium Methods

Ultracentrifugation, Micro-calometry

Titration with competing labelled ligands

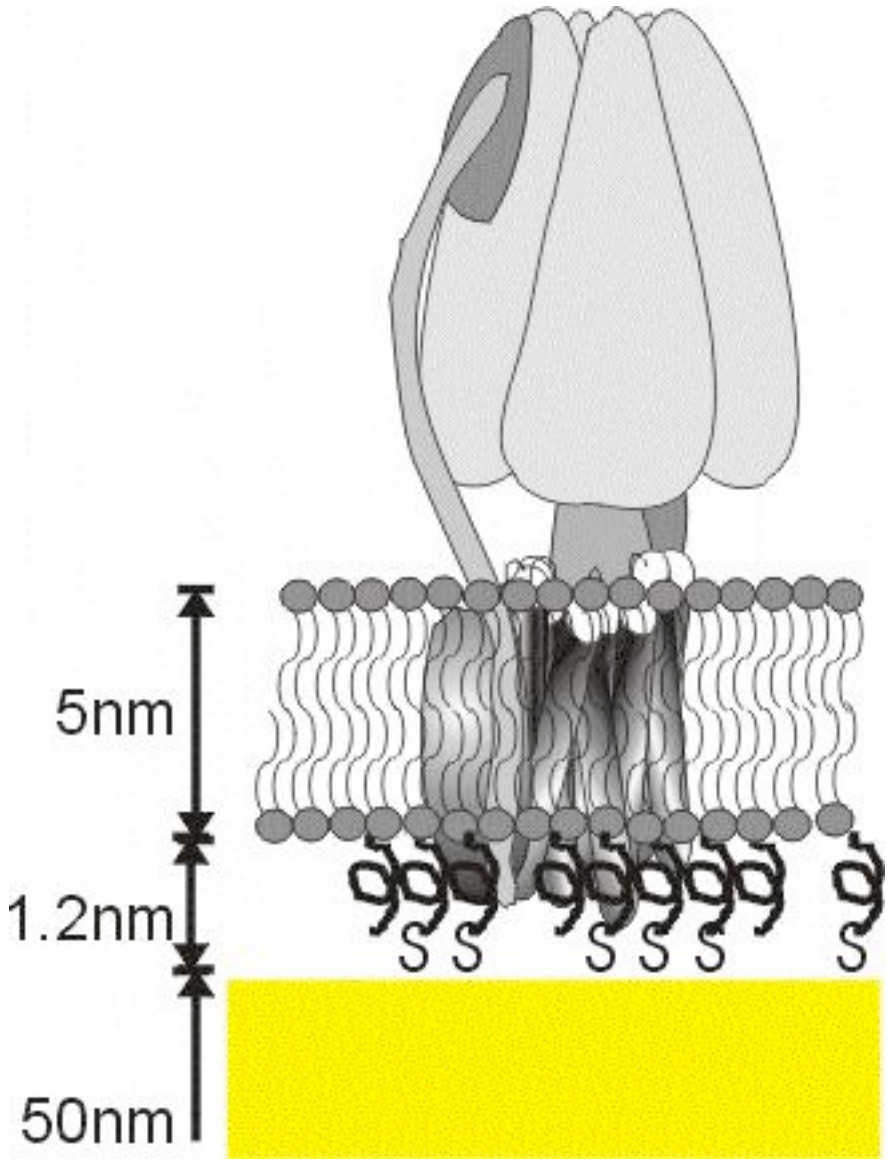
Labelling / Tagging

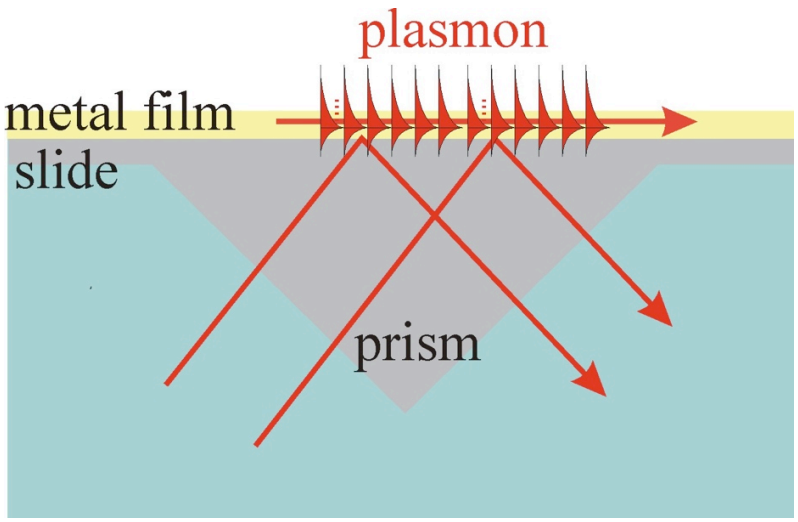
FRET

Two-Hybrid system

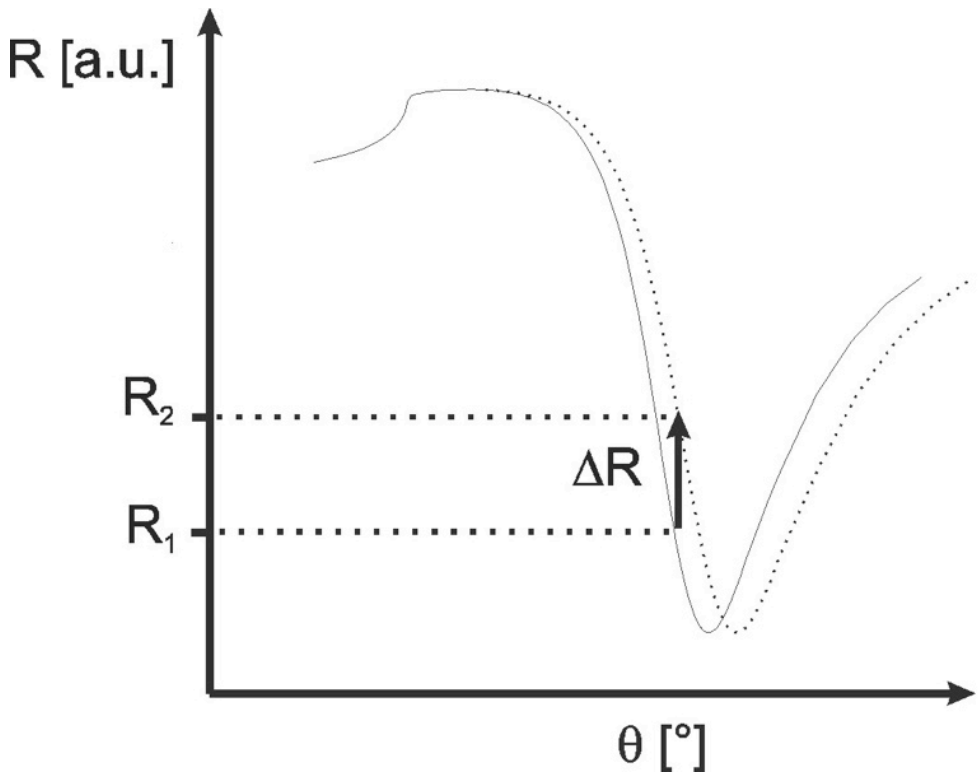
Epitope tagging

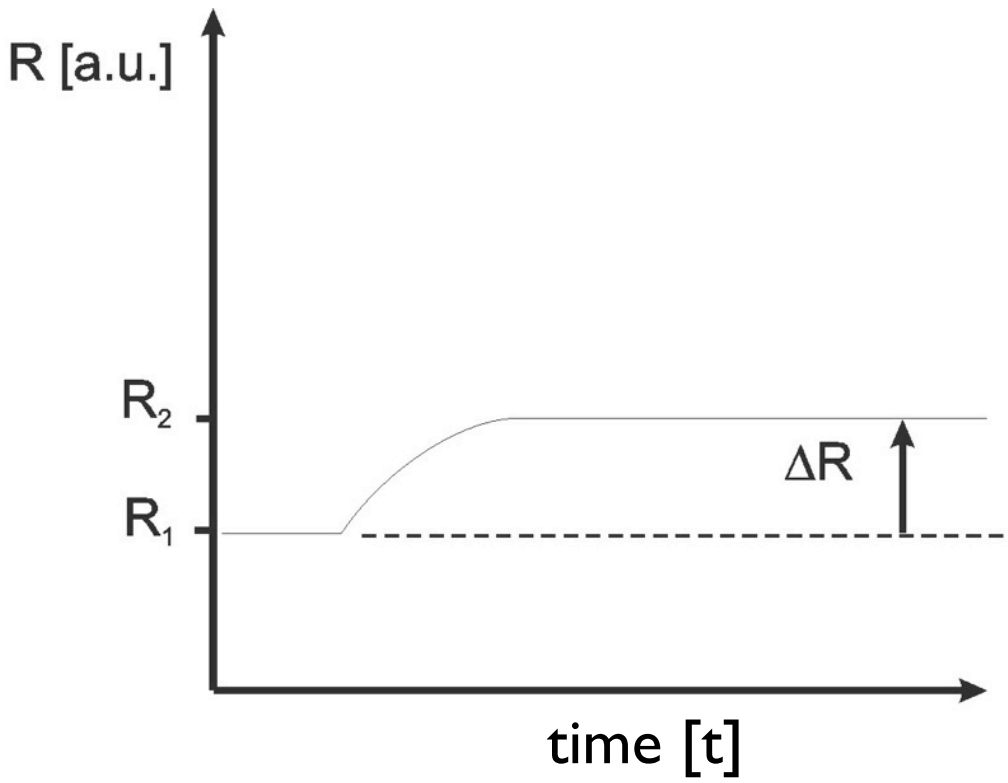
Surface Plasmon Resonance





Surface Plasmon Resonance





Affinity Chromatography and Immunoprecipitation (IP) Allow Identification of Associated Proteins

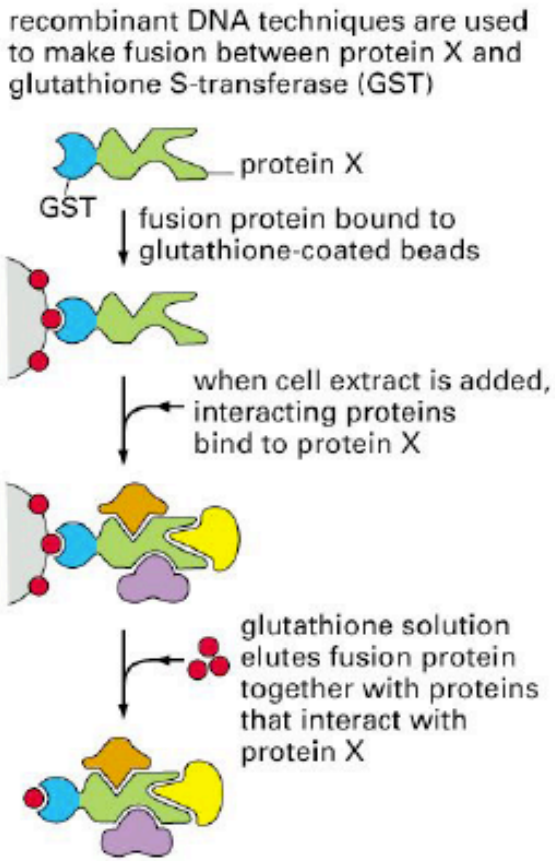


Figure 8–50. Molecular Biology of the Cell, 4th Edition.

Protein affinity chromatography: to isolate and identify proteins that interact physically.

Purification of protein complexes using a GST-tagged fusion protein. Alternatively, a cell extract can be made from a cell producing the GST-fusion protein and pass directly through the glutathione affinity column. The GST-fusion protein, along with proteins associated with it in the cell, are thereby retained.

The identity of these proteins co-purified with GST-protein X can be determined by mass spectrometry.

Co-immunoprecipitation (co-IP): An antibody is used to recognize a specific target protein. Affinity reagents that bind to the antibody and are coupled to a solid matrix are then used to drag the complex out of solution to the bottom of a test tube. Proteins that are associated tightly enough with the target protein also precipitate.

Fluorescence (or Förster) Resonance Energy Transfer (FRET)

The GFP fusion protein strategy has become a standard way to determine the distribution and dynamics of any protein of interest in living cells. GFP, and its derivatives of different colors (such as CFP, RFP and YFP), can also be used to monitor protein-protein interactions using FRET. The most popular FRET pair for biological use is CFP and YFP.

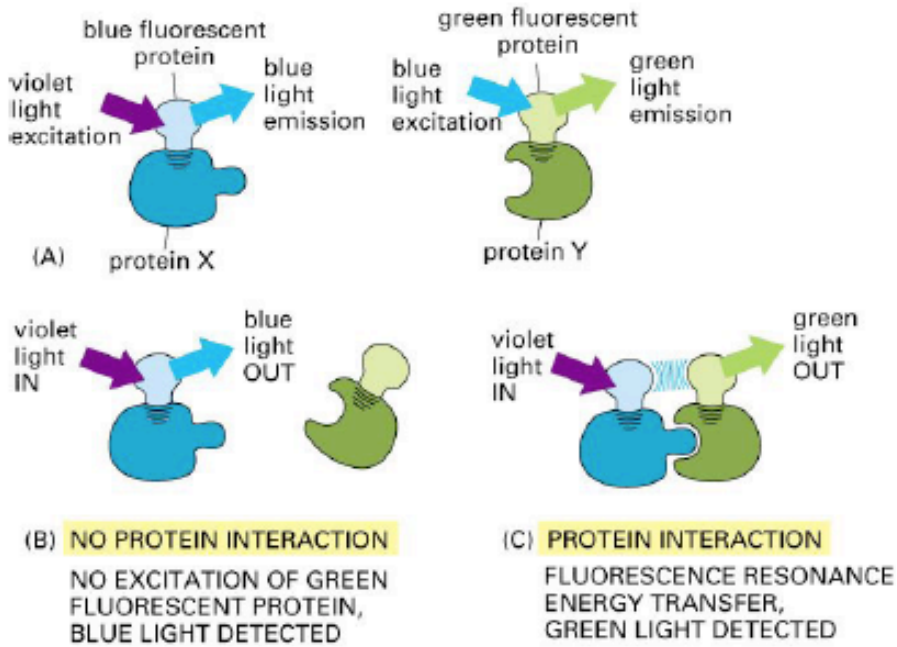


Figure 8–49. Molecular Biology of the Cell, 4th Edition.

To determine whether (and when) two proteins interact inside the cell, the proteins are first produced as fusion proteins attached to different variants of GFP.

(B) If Protein X and Y do not interact, illuminating the sample with violet light yields fluorescence (blue light) from the blue fluorescent protein (X).

(C) If Protein X and Y interact, FRET can occur. Illuminating the sample with violet light excites X , whose emission in turn excites Y, resulting in an emission of green light. The fluorochromes must be quite close (within ~1-10 nm) for FRET to occur.

Limitation: FRET requires external illumination to initiate the fluorescence transfer, which can lead to background noise from direct excitation of the acceptor, or photobleaching. To overcome this difficulty, Bioluminescence Resonance Energy Transfer (or BRET) has been developed. This technique uses a bioluminescent luciferase (typically purified from *Renilla Luciformis*) rather than CFP to produce an initial photon emission compatible with YFP.

*** FRET and BRET are also a common tools in the study of reaction kinetics and molecular motors.**

Protein-Protein Interactions Can Be Identified by Use of the Two-Hybrid System

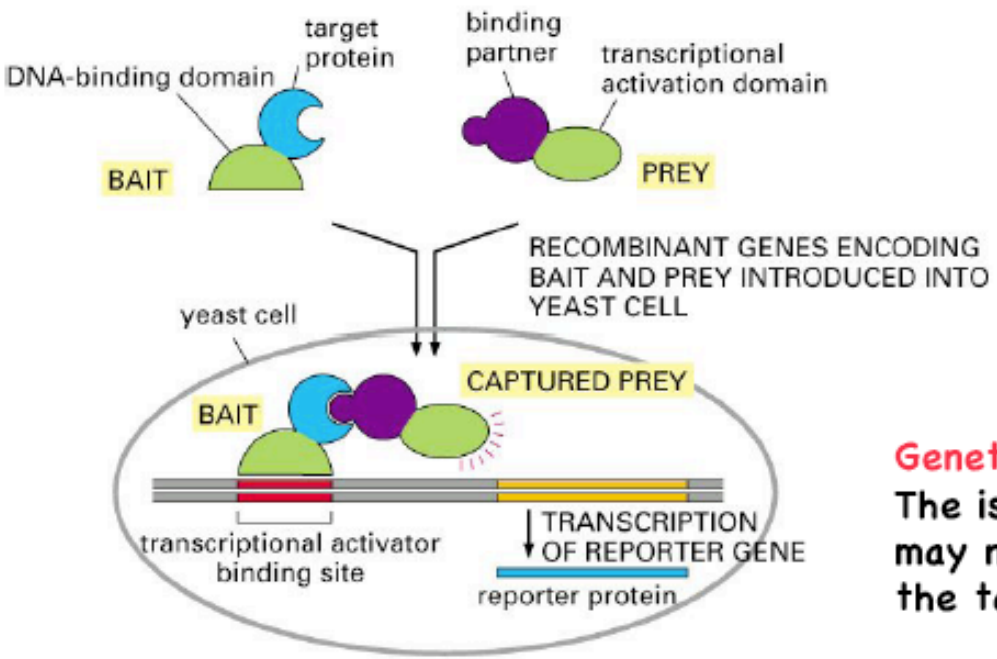


Figure 8–51. Molecular Biology of the Cell, 4th Edition.

Genetic interaction

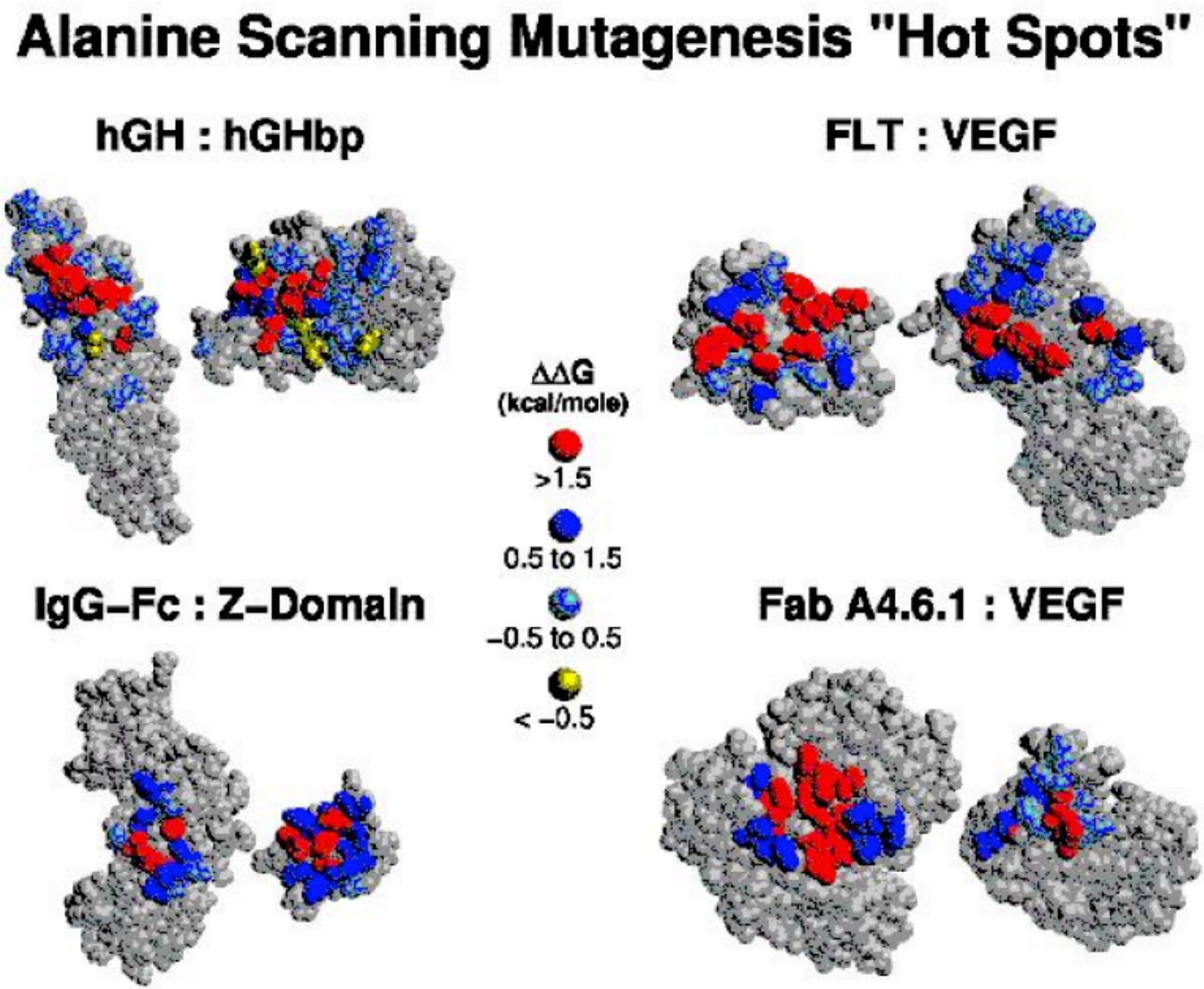
The isolated interacting protein may not directly interact with the target protein.

Two-hybrid systems allow the simultaneous isolation of interacting proteins along with the genes that encode them.

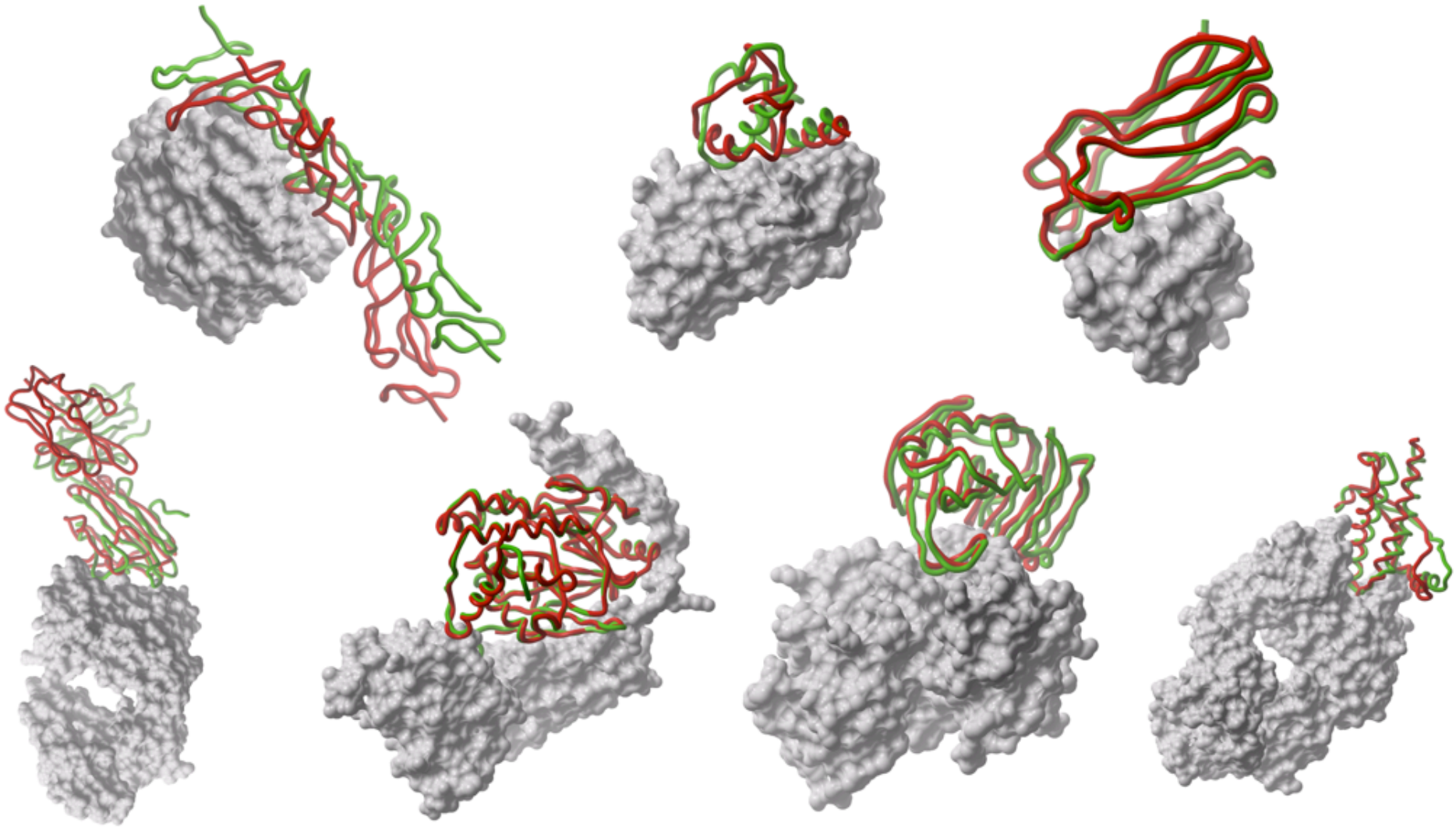
Of the thousands of protein-protein interactions that have been catalogued in yeast, half have been discovered with such 2-H screens.

2. Protein-protein interactions hotspots

- Alanine Scanning Mutagenesis



Protein Docking : 6 Dimensions and More



http://mmb.pcb.ub.es

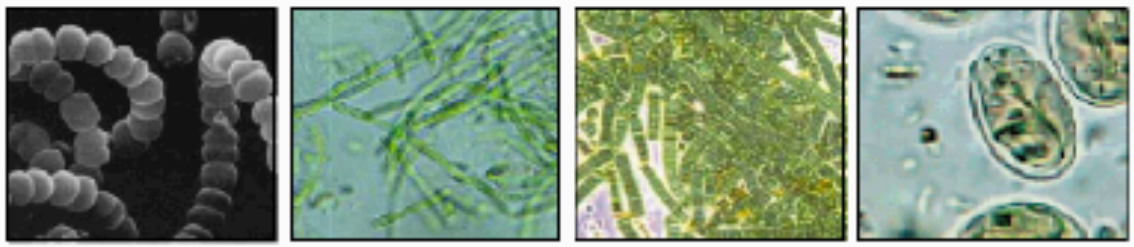
Computational Methods for predicting PPI

- Phylogenetic Profiles
- Rosetta Stone
- Gene Neighbors
- Co-evolution
- Gene clusters

Phylogenetic Profile

Figure form Bowers at al., Genome Biology 2004

(c)



	Genome 1	Genome 2	Genome 3	Genome 4
Protein A	1	1	1	0
Protein B	1	1	1	0
Protein C	1	0	1	1
Protein D	1	1	0	1

The same profiles

- Idea: Pairs of non-homologous proteins that are always both present or both absent in a genome suggest their functional dependence → possible interaction
- Profile of a protein: A vector of 0/1 where each position corresponds to one genome: 1- protein present 0-protein absent

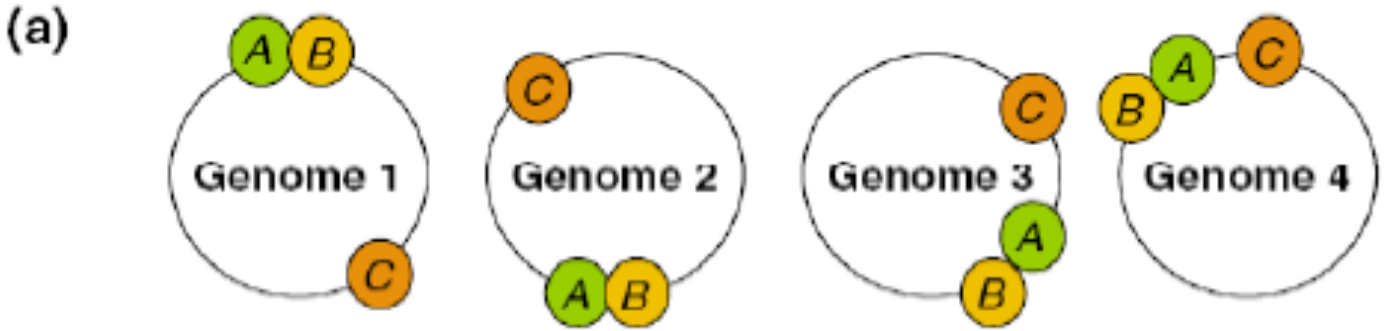
Gene Cluster Method

Within bacteria, proteins of closely related function are often transcribed from a single functional unit known as an operon. Operons contain two or more closely spaced genes located on the same DNA strand. These genes are often in proximity to a transcriptional promoter that regulates operon expression.

Advantage: Each operon is informative (multiple genome comparison is not necessary)

Gene Neighbors

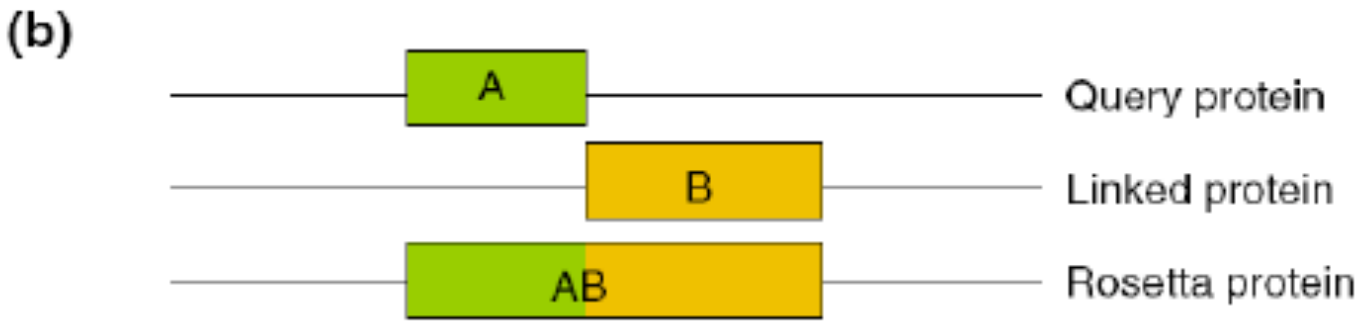
Figure form Bowers et al, Genome Biology 2004



- Gene A is a neighbor of B in several genomes - potential functional link

Rosetta Stone

Figure form Bowers, Genome Biology 2004

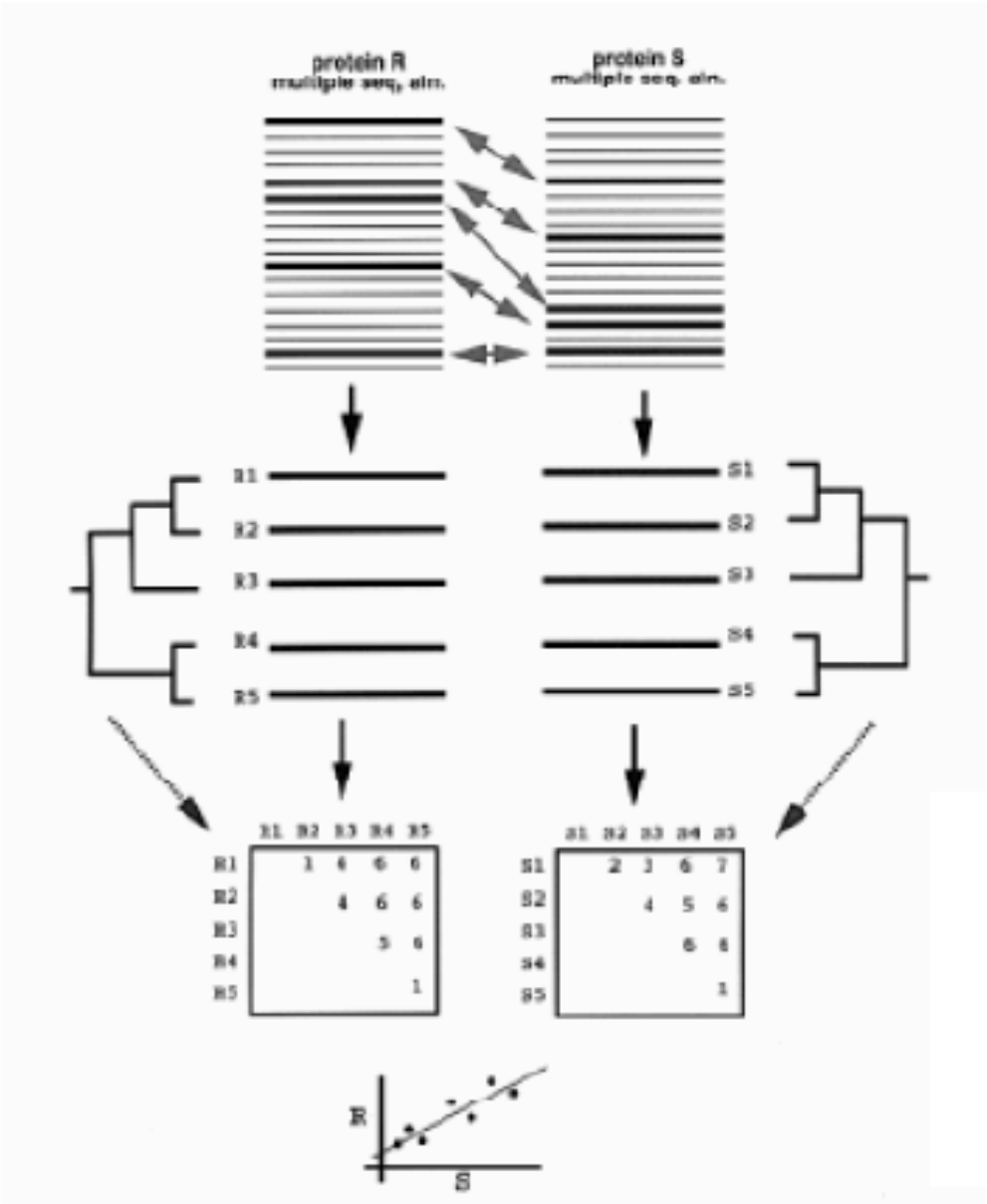


- A, B – two domains
- Rosetta protein – protein containing both domains in some organism – indication that in anther organism these two domains (which now are in different proteins) may interact

Co-evolution method

- Idea: Assume that protein A and B interact.
- If A and B are both present in several organisms and perform the same role in these organisms they interact in all these organisms
- Evolution of A and B should be correlated

Comparing the trees



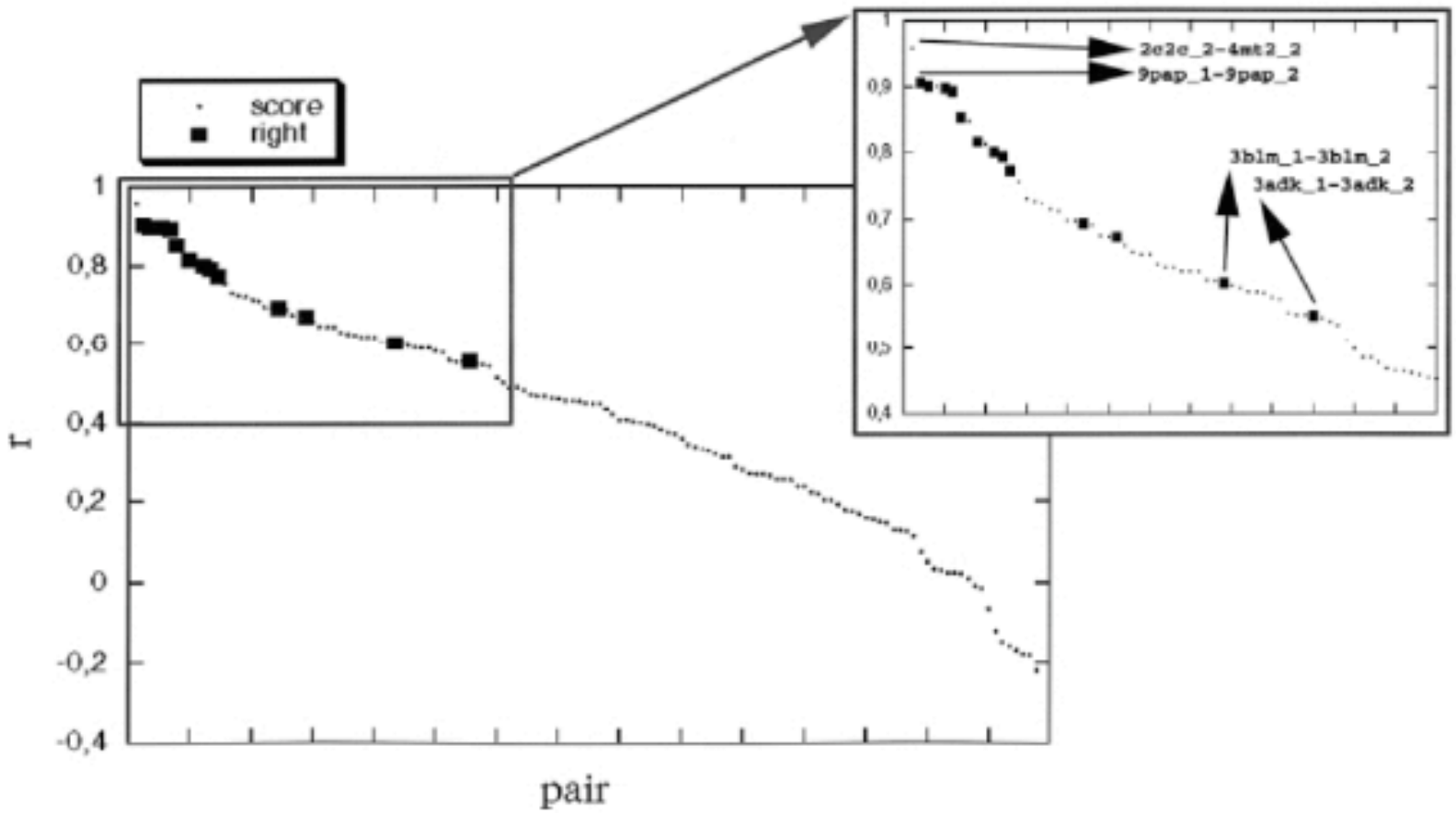
Rather than computing similarity of two trees compute correlation coefficient of the two distance matrices

$$r = \frac{\sum_{i=1}^{n} (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n} (R_i - \bar{R})^2} \sqrt{\sum_{i=1}^{n} (S_i - \bar{S})^2}}$$

Results

Distance matrices for families of two interacting proteins are have high correlation coefficient

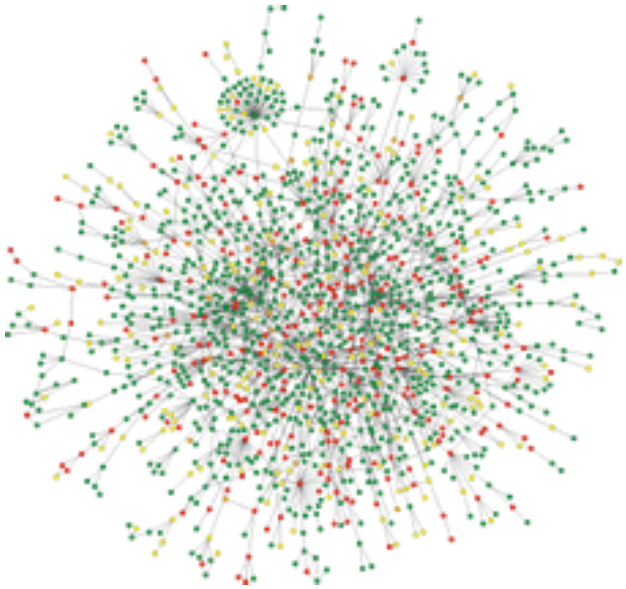
F.Pazos and A.Valencia

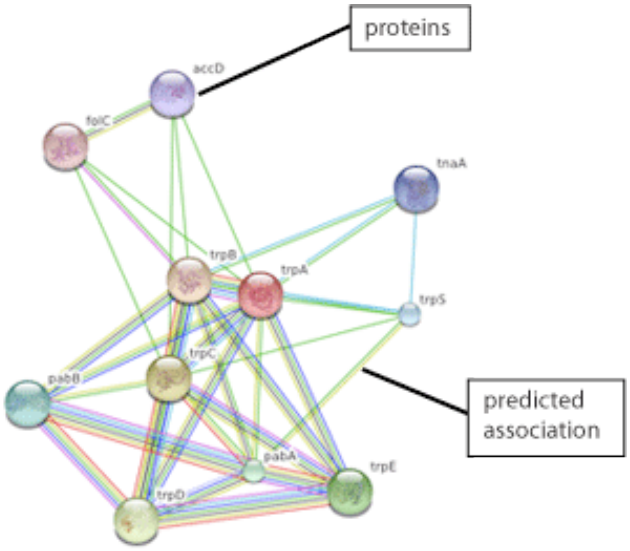


PROTEIN INTERACTION NETWORKS

Scale free

Power law distribution (?)

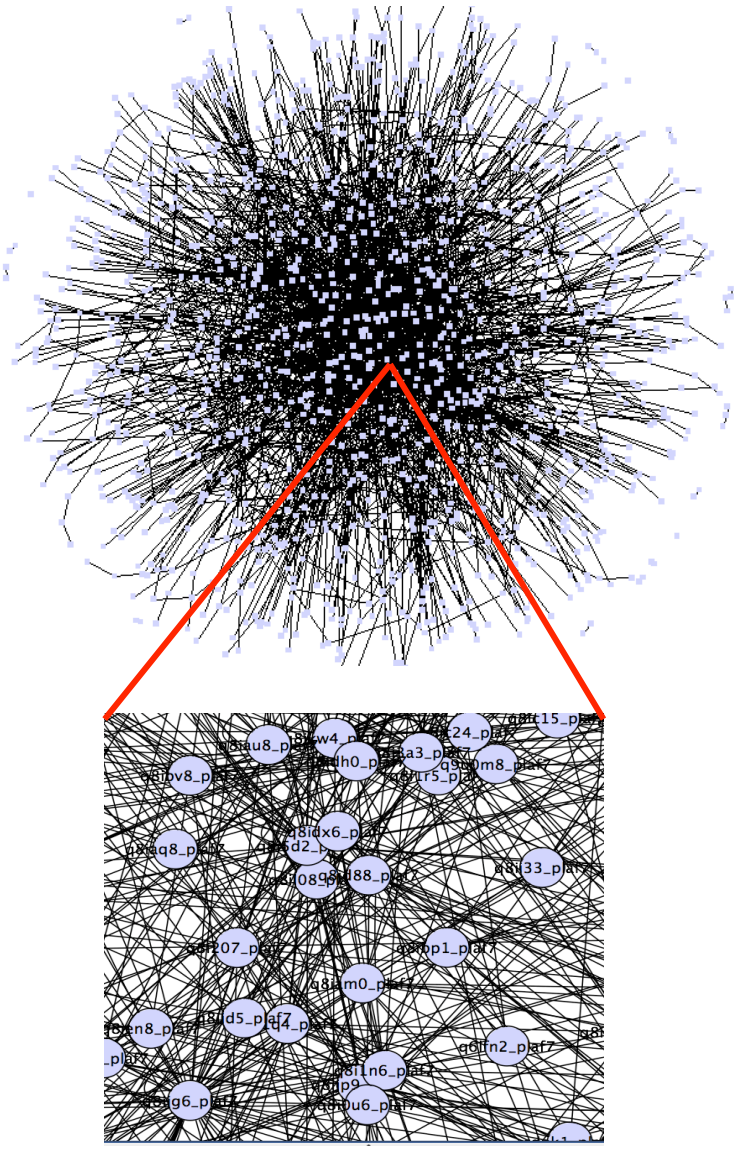




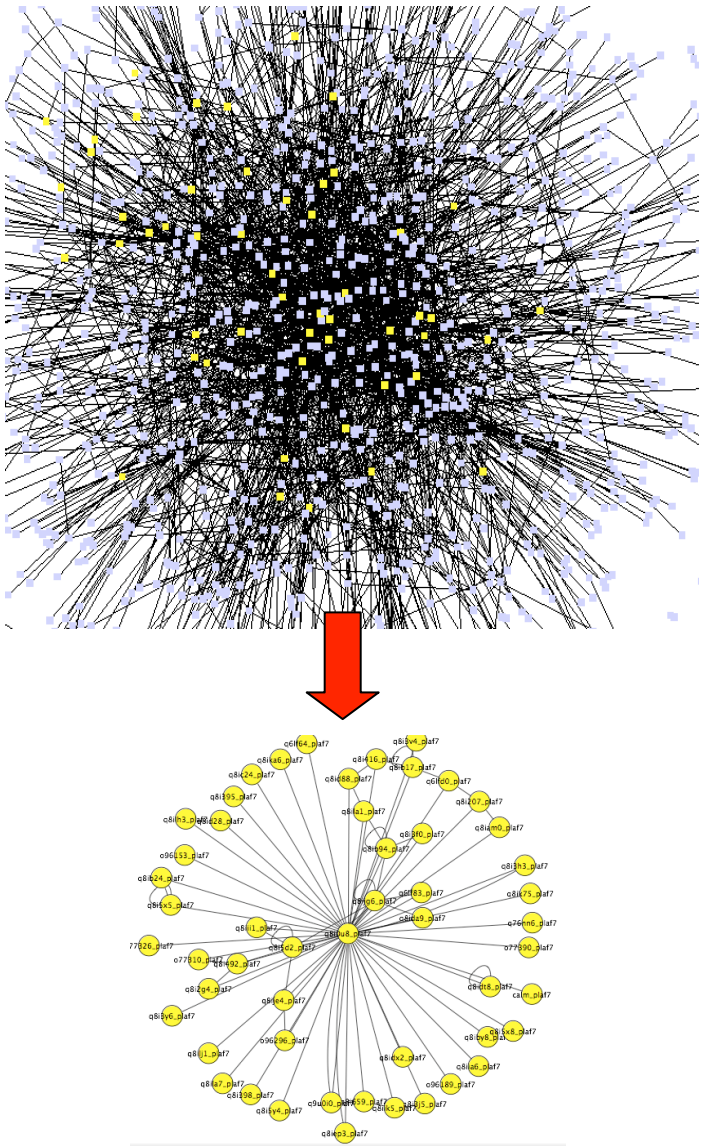
Undirected graphs

Hub Proteins

Plasmodium Falciparum Network



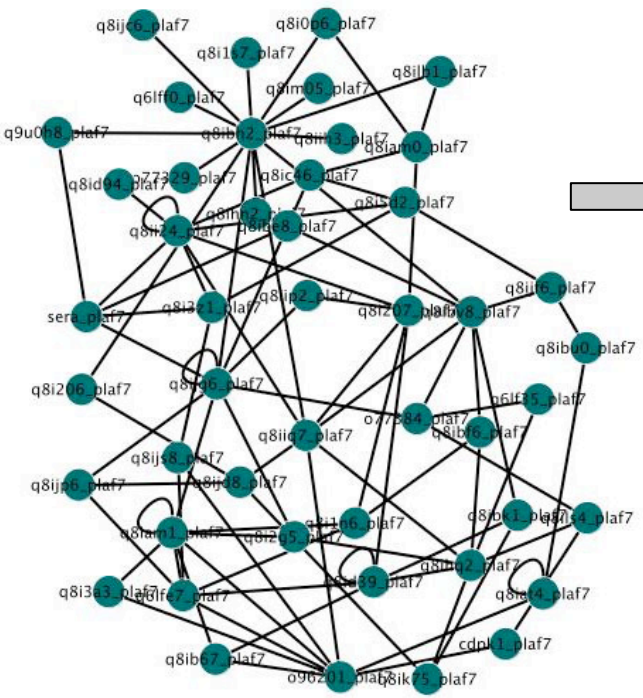
MSP-1 Protein Neighbors



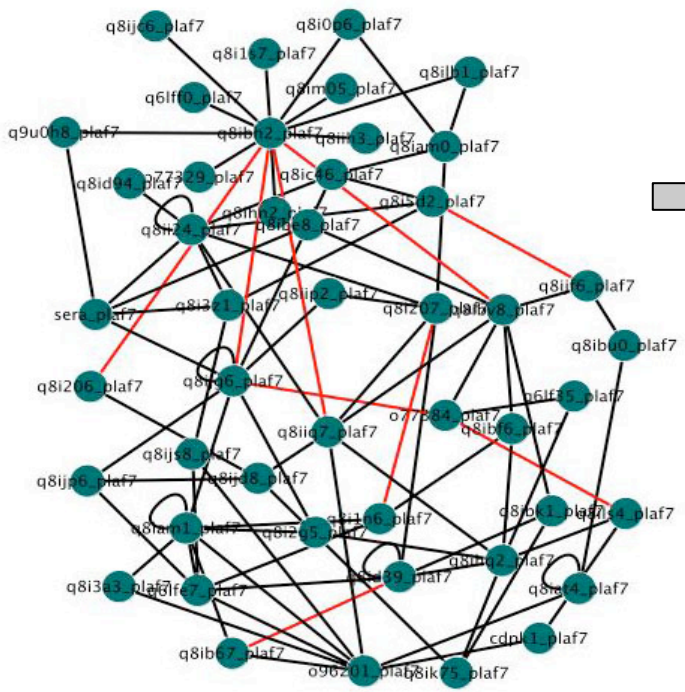
- Hub Proteins are essential nodes in the networks, when attacked they are more lethal to the network topology than the average protein.
- They have above average partners and should represent the most essential parts of the interactome.
- Two classes of hubs have been proposed “date hubs” and “party hubs”, the first interact with each partner at different times while the second interacts with most partners at the same time

Analysis of Structure interaction Network:

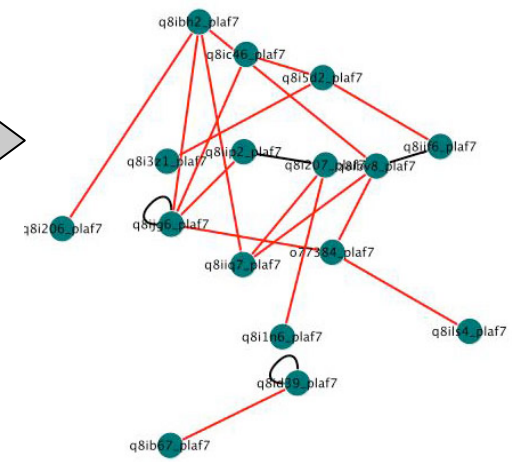
P-P Interactions Network



Filtered 3D Interactions



Structure Interaction Network



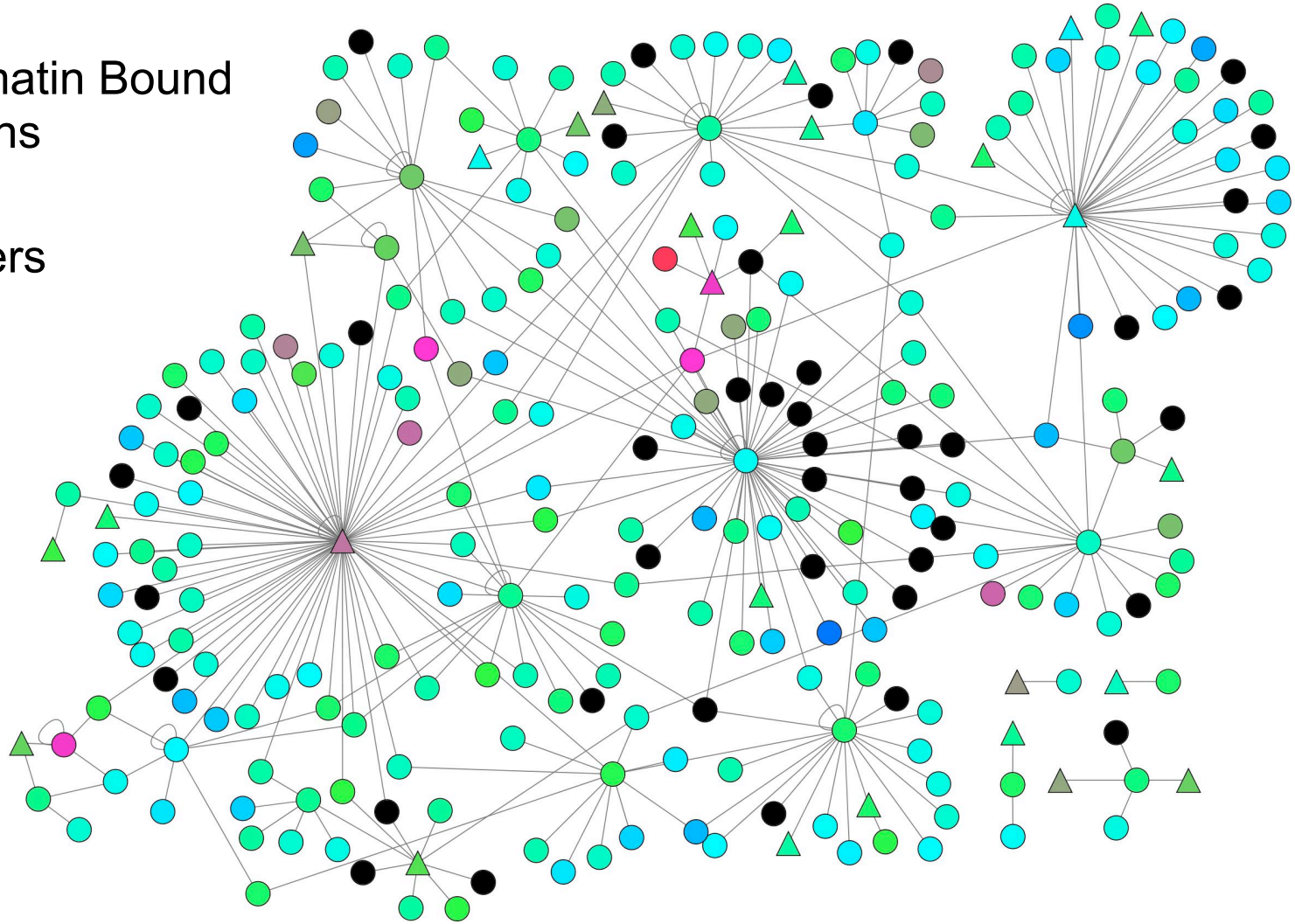
Kim PM, Lu LJ, Xia Y, Gerstein MB. Science. 2006 314(5807):1938-41

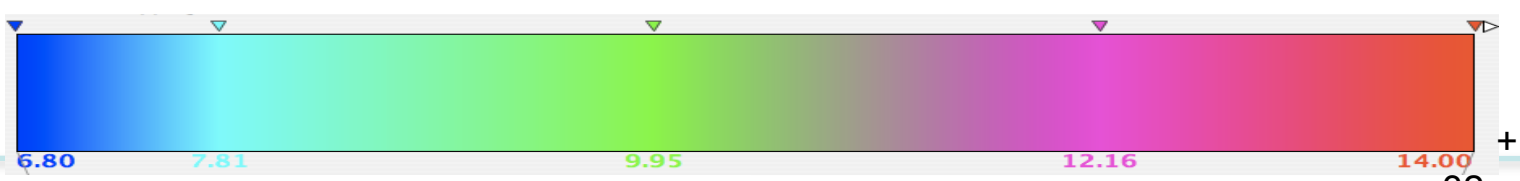
T-cells Cell Cycle SNK: Δ0h-4h Network

Chromatin Bound Proteins

Partners

t = 4h





- low expression levels

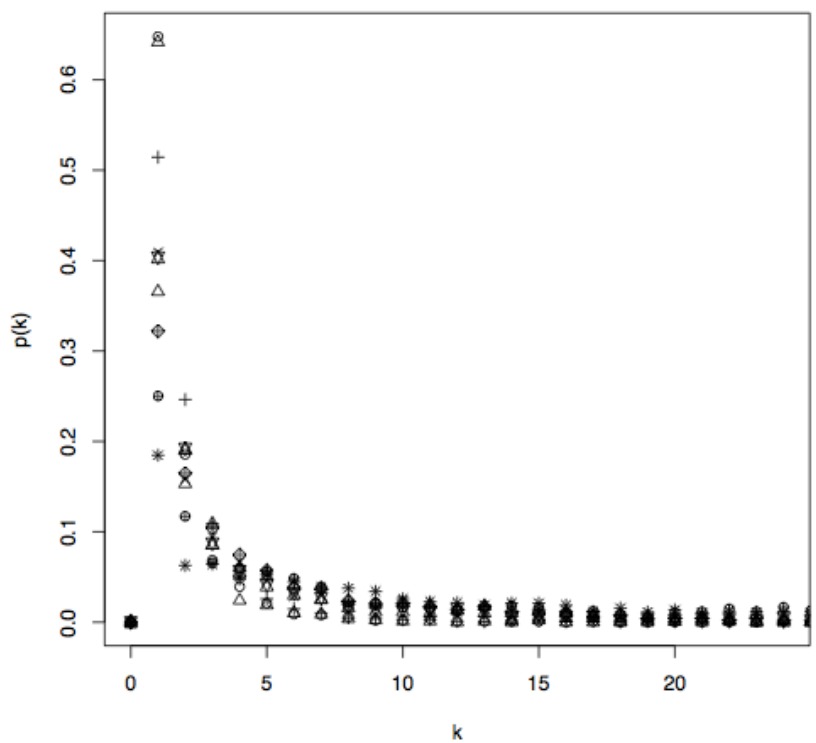
+ high expression levels

Interaction Matrices

A standard way of characterizing the topology of a network $\mathbf{c}$, as e.g. observed in a biological or physical system under study, is to measure for each vertex i the degree $k_i(\mathbf{c}) = \sum_j c_{ij}$, the number of links to this vertex. From these numbers then follow the empirical degree distribution $p(k|\mathbf{c})$ and the observed average connectivity $\bar{k}(\mathbf{c})$:

$$p(k|\mathbf{c}) = \frac{1}{N}\sum_i \delta_{k,k_i(\mathbf{c})}, \qquad \bar{k}(\mathbf{c}) = \frac{1}{N}\sum_i k_i(\mathbf{c}) \tag{1}$$

degree distribution



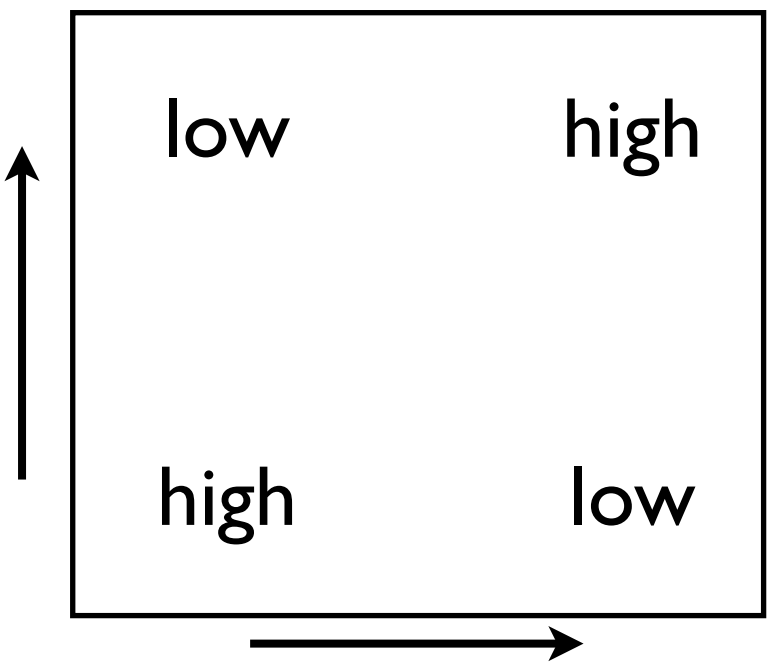
S. cerevisiae

Interaction Matrices

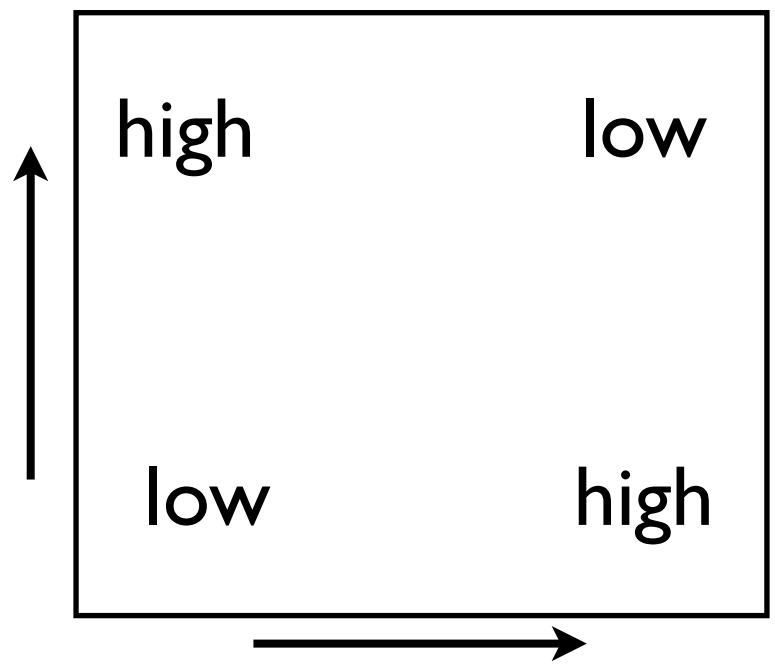
$$a = \frac{\langle kk' \rangle - \langle k \rangle^2}{\langle k^2 \rangle - \langle k \rangle^2}$$

assortativity

positive



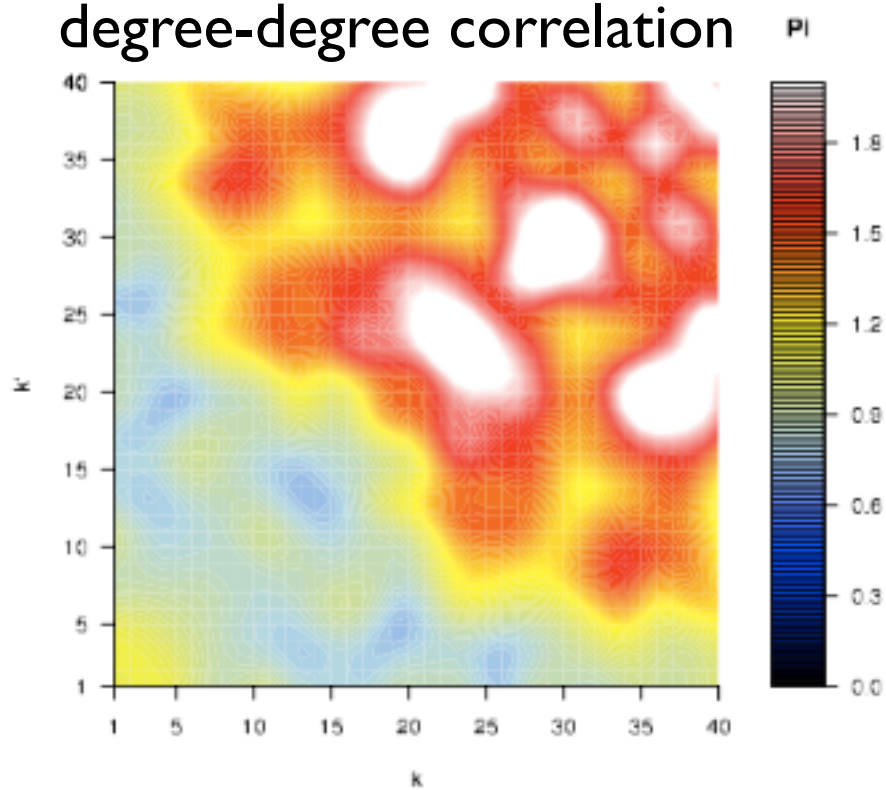
negative



Interaction Matrices

$$\Pi(k,k'|\mathbf{c}) = \frac{\tilde{\Pi}(k,k'|\mathbf{c})}{\tilde{\Pi}(k,k'|\mathbf{c}_0)} = \frac{\mathcal{P}[\mathrm{conn}|\mathbf{c},k,k']}{\mathcal{P}[\mathrm{conn}|\mathbf{c}]}\frac{\overline{k}^2(\mathbf{c})}{kk'}$$

degree-degree correlation



E.coli

Learning Outcomes

- Obligate / Non-obligate Interactions
- Spectrum of Protein Interactions
- Allostery
- Thermodynamics / Kinetics of Interaction Reaction
- Experimental Methods for Complexes
- Computational Methods for Complexes
- Interaction Networks