Analysis and Prediction of Protein Complexes

Proteomes Interactomes and Biological Networks

Emidio Capriotti

http://biofold.org/

Biomolecules Folding and Disease

Department of Pharmacy and Biotechnology (FaBiT)
University of Bologna

The Molecular Viewpoint

- The affinity of PPI varies from millimolar to picomolar, depending on the type of interaction and signaling needed (Chen et al. Protein Sci. 2013)

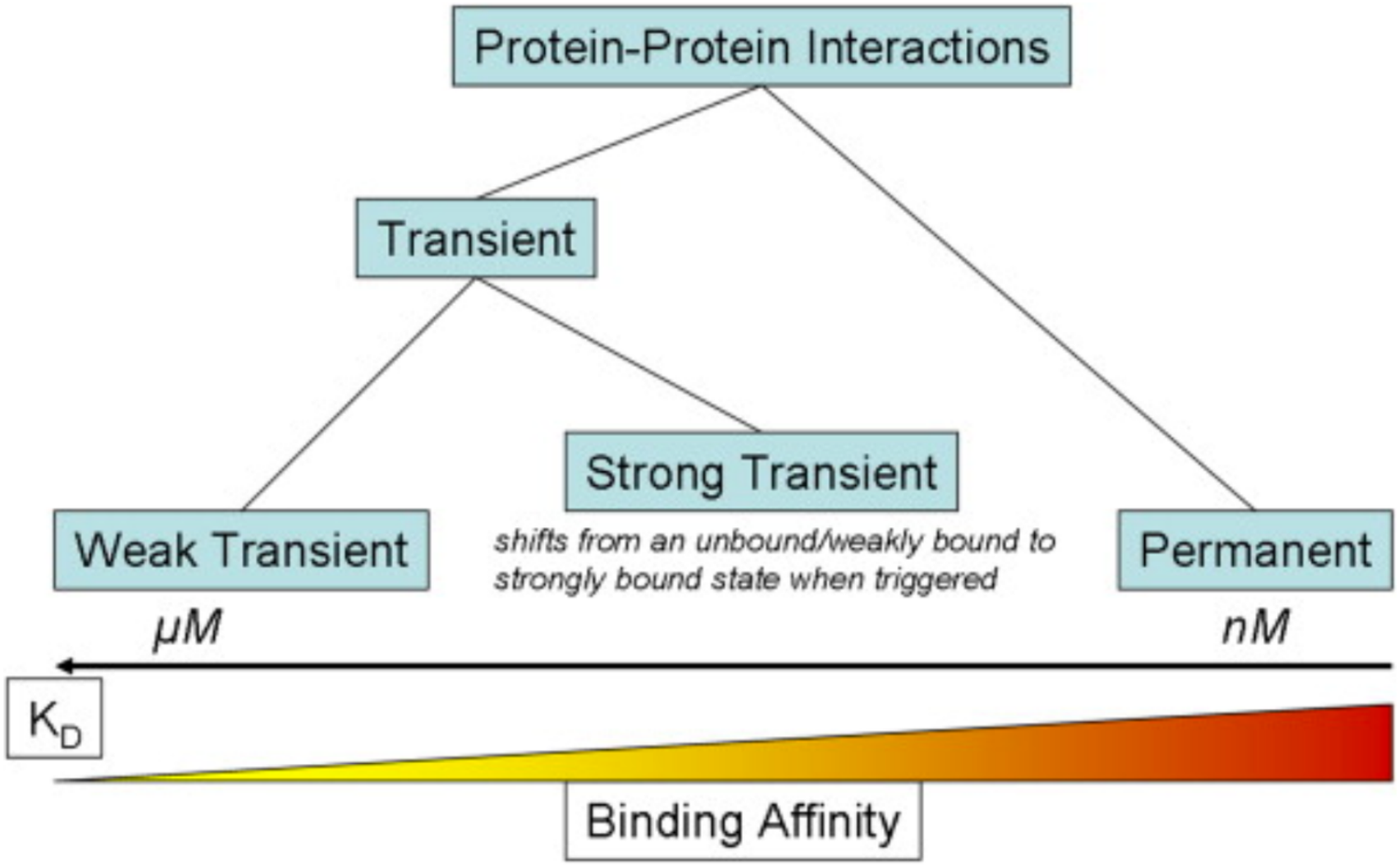
- Despite affinity varies over a wide range, proteins maintain a high degree of specificity for their partners

- Many proteins exhibit specificity for multiple partners (Reichmann et al. Curr. Opin. Struct. Biol. 2007).

- The nature of the interaction surface determines how proteins interact

- A detailed knowledge of the interaction surfaces of proteins and their energetics is necessary to understand the regulatory mechanisms of biochemical pathways (especially to modulate or block these pathways for therapeutic purposes)

Protein-Protein Interactions



Strong transient: This category includes interactions that are triggered/stabilised by an effector molecule or conformational change. An example is given by the Ras proteins, which form tight complexes with their partners when GTP-bound and only weak complexes when GDP-bound.

Perkins et al, Structure, 2010

Surface of Interaction (I)

- The area of PPI interfaces is large (1000 to 4000 Å^2)
- Standard-sized interfaces are 1200 to 2000 Å^2
- Short-lived and low-stability complexes -> smaller interfaces (1150–1200 Å^2)
- large surfaces (2000 to 4600 Å^2) ->
  - proteases and particular inhibitors
  - G-proteins and other components of the signal transduction system
- Protein-small molecule interaction surfaces have an area of 300 to 1000 Å^2.

Surface of Interaction (II)

- Surfaces of PPIs are generally flat and lack the grooves and pockets that are present at the surfaces of proteins that bind to small molecules.

- PPI surfaces are generally hydrophobic in nature.

- Only certain hydrophobic spots contribute to the free energy of binding and help to hold the two proteins together.

- Such regions are called hot spots.

Hot Spots

- Hot spots account for less than 50% of the contact area of PPI
- A region of protein surface is called a hot spot when replacement of an amino acid residue by alanine in that spot lowers the free energy of binding by at least 2 kcal/mol
- Analysis of the amino acid composition of hot spots shows that some residues are found more frequently in hot spots (Tyr, Trp, and Arg)
- The hot spots are surrounded by energetically less important residues that separate/prevent bulk water from hot spots

Analysis of Protein Complex

- identification of interface residues/hot spots

- details about the interface
  solvent accessible surface area, shape, complementarity between surfaces, residue interface propensities, hydrophobicity, segmentation and secondary structure, and conformational changes on complex formation

- assignment of protein function

- recognition of specific residue motifs

Structure PPI Data

- The most significant contribution to understanding the PPI surface comes from structural biology via X-ray crystallography or NMR as well as mutational studies

- Prediction of interaction/binding sites

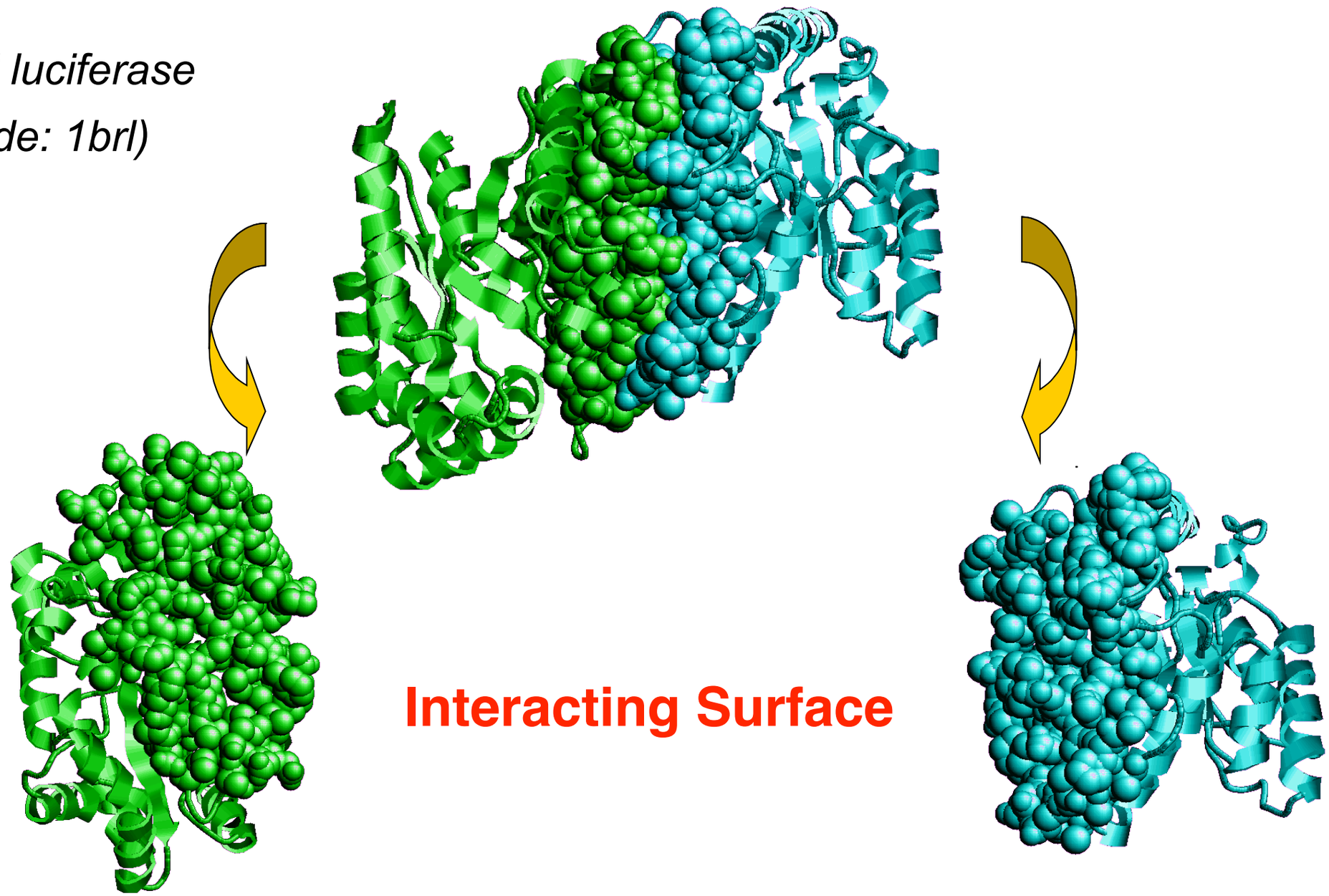
- Prediction of protein-protein complexes

Interacting surface

Difference in Accessible Surface Area (ASA) between monomers and complex

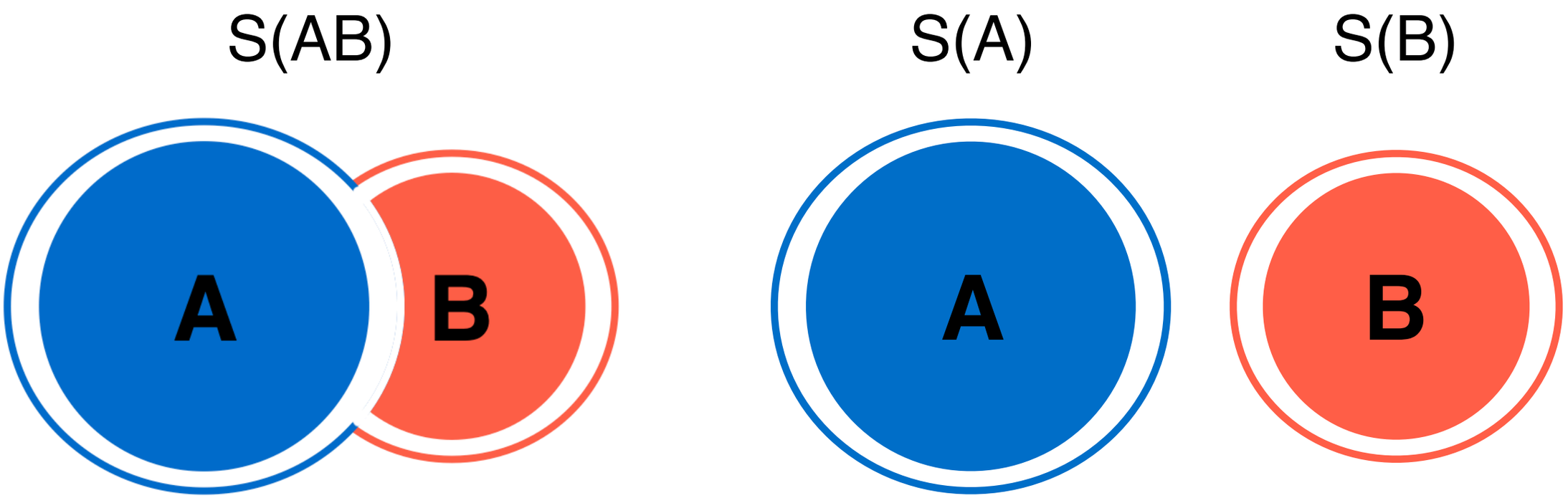
Bacterial luciferase

(PDB code: 1brl)



Protein dimer

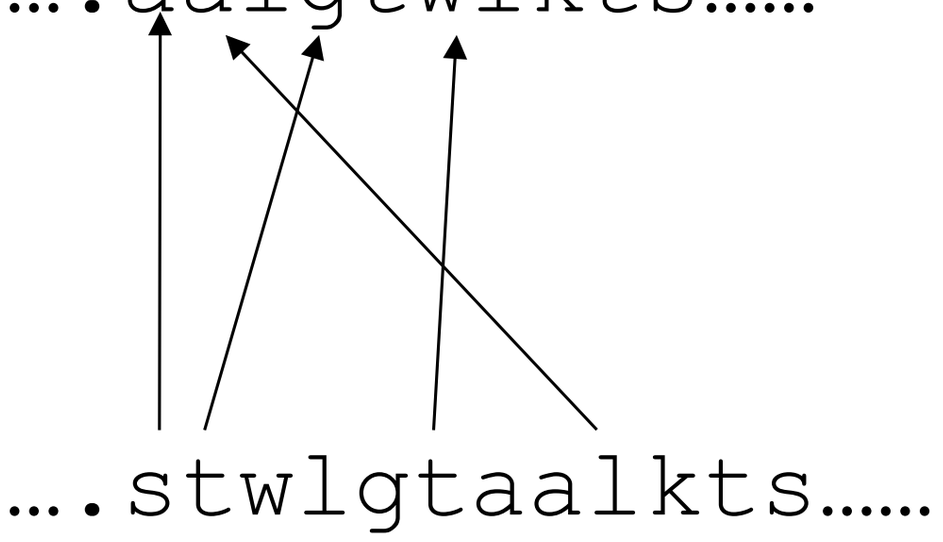
For a protein dimer the interacting surface can be calculated as follows:



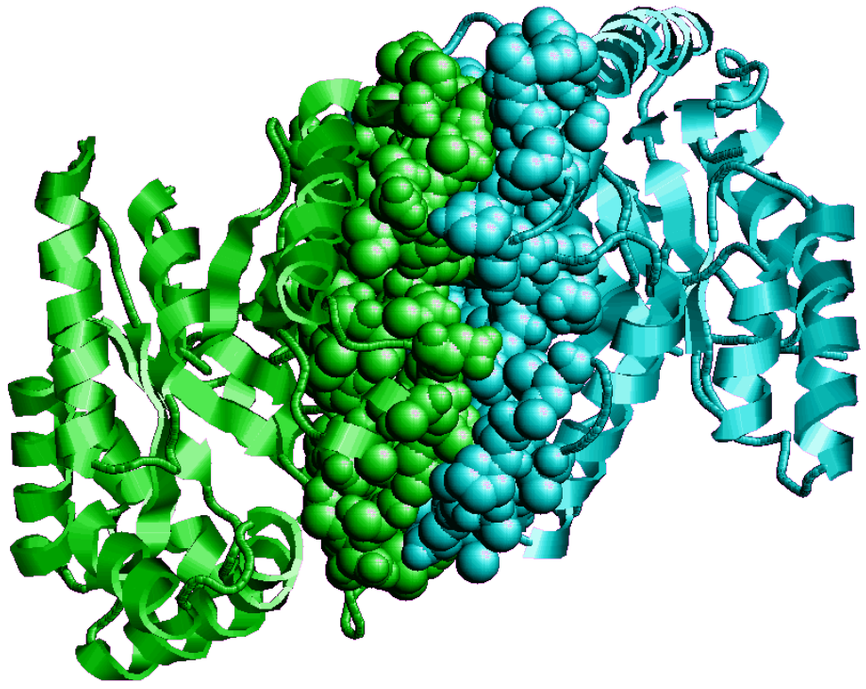
$$\text{Interacting Surface (AB)} = \frac{S(A) + S(B) - S(AB)}{2}$$

Prediction features

Protein Sequence



Protein Structure



+ Whole genome computation

- No exact location, No atomic description

+ Exact location Atomic description

- Availability of the 3D coordinates

Three major problems

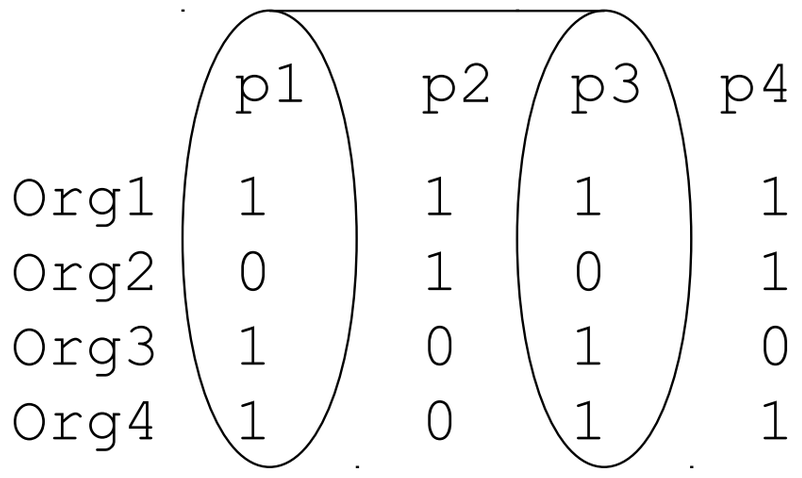
- Protein-Protein interaction networks: given a set of proteins, predict the possible partners

- Docking: given a pairs of proteins, known to interact, predict the geometry of the complex

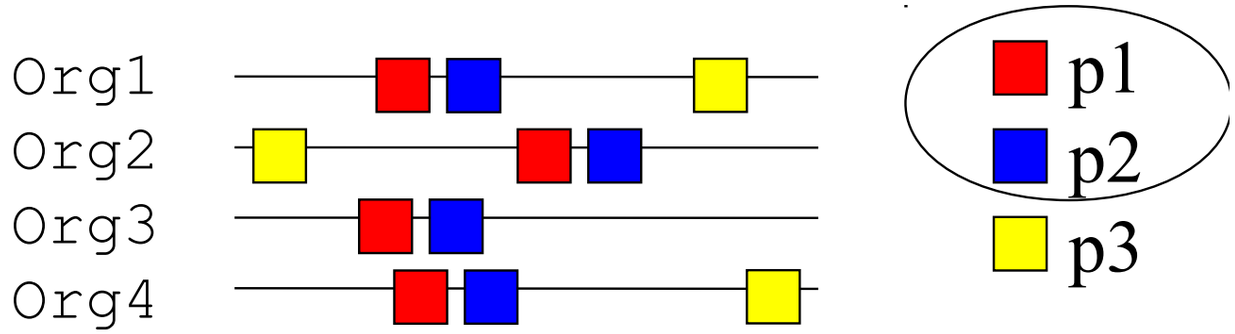
- Protein-interaction sites: given a single protein, predict possible interacting regions

Sequence-based methods

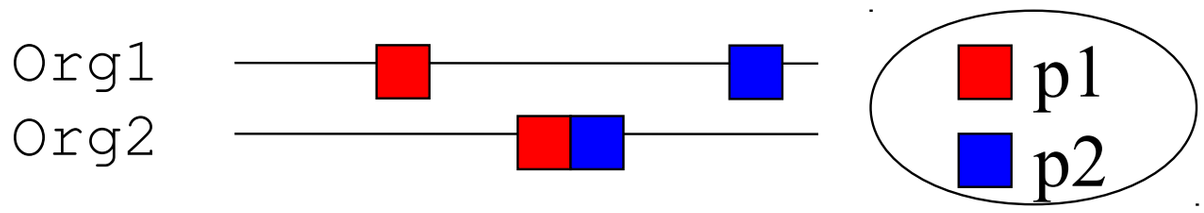
Phylogenetic Profiling: interacting proteins should co-evolve and should have orthologs in closely related species.



Gene Neighborhood: interacting proteins and co-evolving homologs tend to have close genomic locations.



Gene Fusion: two proteins that interact may have homologs in other genomes that are fused into a unique protein

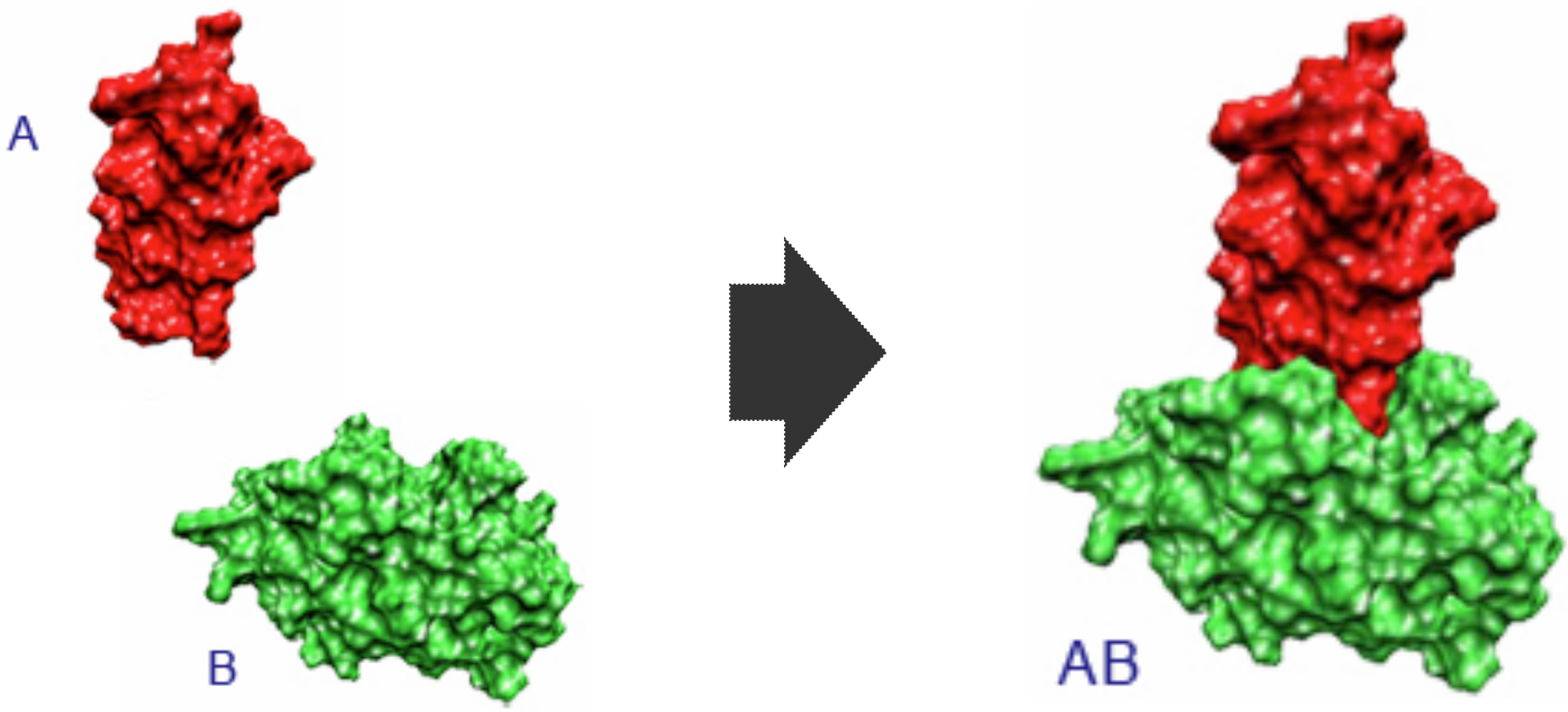


From wikipedia

Protein Docking

- Computational schemes that aims to **find the "best" matching between two molecules**, a **receptor** and a **ligand**

- The molecular docking problem can be defined as follows: **given the atomic coordinates of two molecules, predict their "correct" bound association**



Halperin et al. Proteins, 2002

Some methods

- **HADDOCK** (software/web server).
  http://haddock.chem.uu.nl

- **CLUSPRO** (software/web server)
  http://cluspro.bu.edu

- **ICM-pro** (desktop-modeling environment)
  http://www.molsoft.com/protein_protein_docking.html

- **ROSETTADOCK** (software/web server)
  http://graylab.jhu.edu/docking/rosetta/
- http://rosettadock.graylab.jhu.edu/submit

- **GRAMM-X** (web server)
  http://vakser.bioinformatics.ku.edu/resources/gramm/grammx

- **PATCHDOCK/FIREDOCK** (software/web server)
  http://bioinfo3d.cs.tau.ac.il/PatchDock/

- **HEX** (software/web server)
  http://hexserver.loria.fr

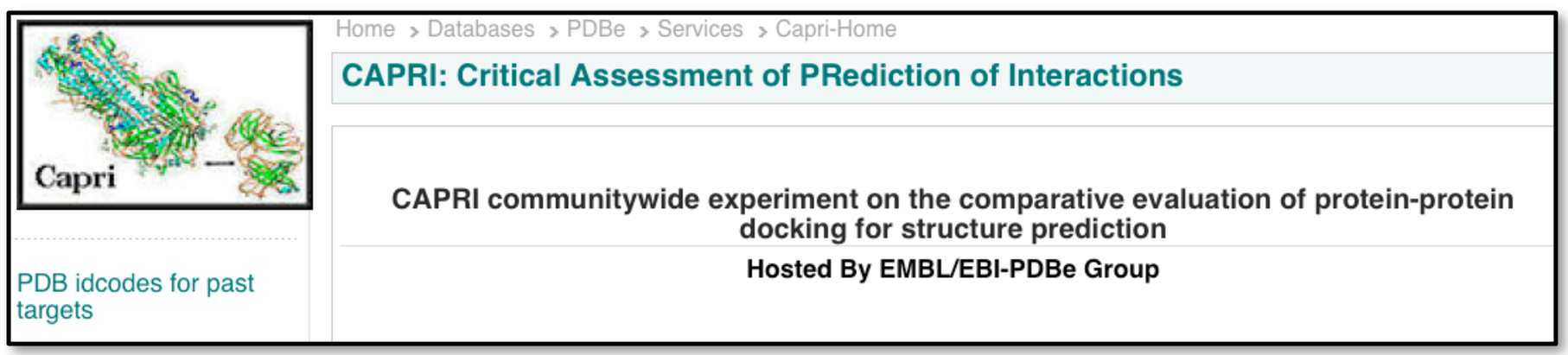
Docking limitations

Although the docking algorithms shows good performances in the prediction of enzyme-inhibitor and antigen-antibody complexes. The molecular docking problem is far from being solved

- It is difficult to find very specific properties of protein-protein interfaces
- Results are generally poor with weakily interacting proteins
- Proteins are flexible and may undergo even large conformational changes upon binding

CAPRI Experiments

- CAPRI is a community-wide experiment in modelling the molecular structure of protein complexes
- CAPRI is a blind prediction experiment aimed at testing the performance of protein docking methods
- Rounds take place about every six months
- Each round contains between one and six target protein–protein complexes whose structures have been recently determined experimentally
- Targets are unpublished crystal or NMR structures of complexes, whose coordinates are held privately by the assessors, with the co-operation of the structural biologists who determined them
- The atomic coordinates of the two proteins are given to groups for prediction



http://www.ebi.ac.uk/msd-srv/capri/

Exercise

Download the PDB file of the Bacterial luciferase (Vibrio harveyi) from the PDB (code: 1BRL)

- Generate the DSSP file for the protein complex and the isolated chains A and B
- Calculate the total solvent accessible area of the complex and isolated chains and calculate the surface of interaction for both chains.
- Given the size of the binding surface what kind of protein interaction it is expected?
- Find the residue at the interface and calculate the variation of relative solvent accessible area. Which residue are buried in the interacting surface?

Chain = col 12, AA = col 14, SS = col 17, Acc: cols 36-38, Phi: cols 104-109, Psi: cols 110-115