



# Protein interactions and binding site properties

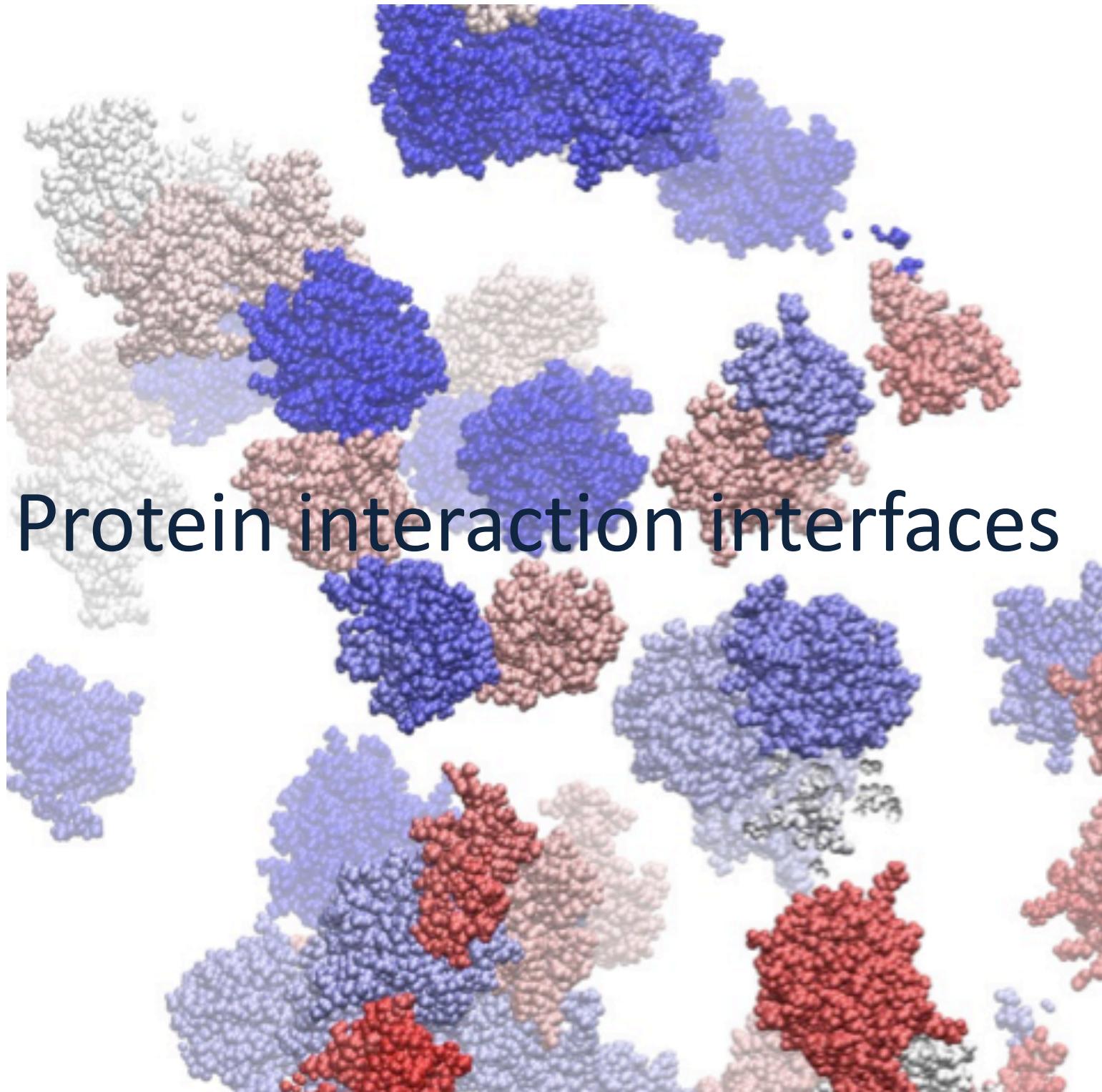
Allegra Via



Consiglio Nazionale delle Ricerche



<https://goo.gl/forms/O9KA9hE8azNQAKLD3>



# Protein interaction interfaces

## **A) Protein-ligand interactions**

- Protein interactions and binding site properties

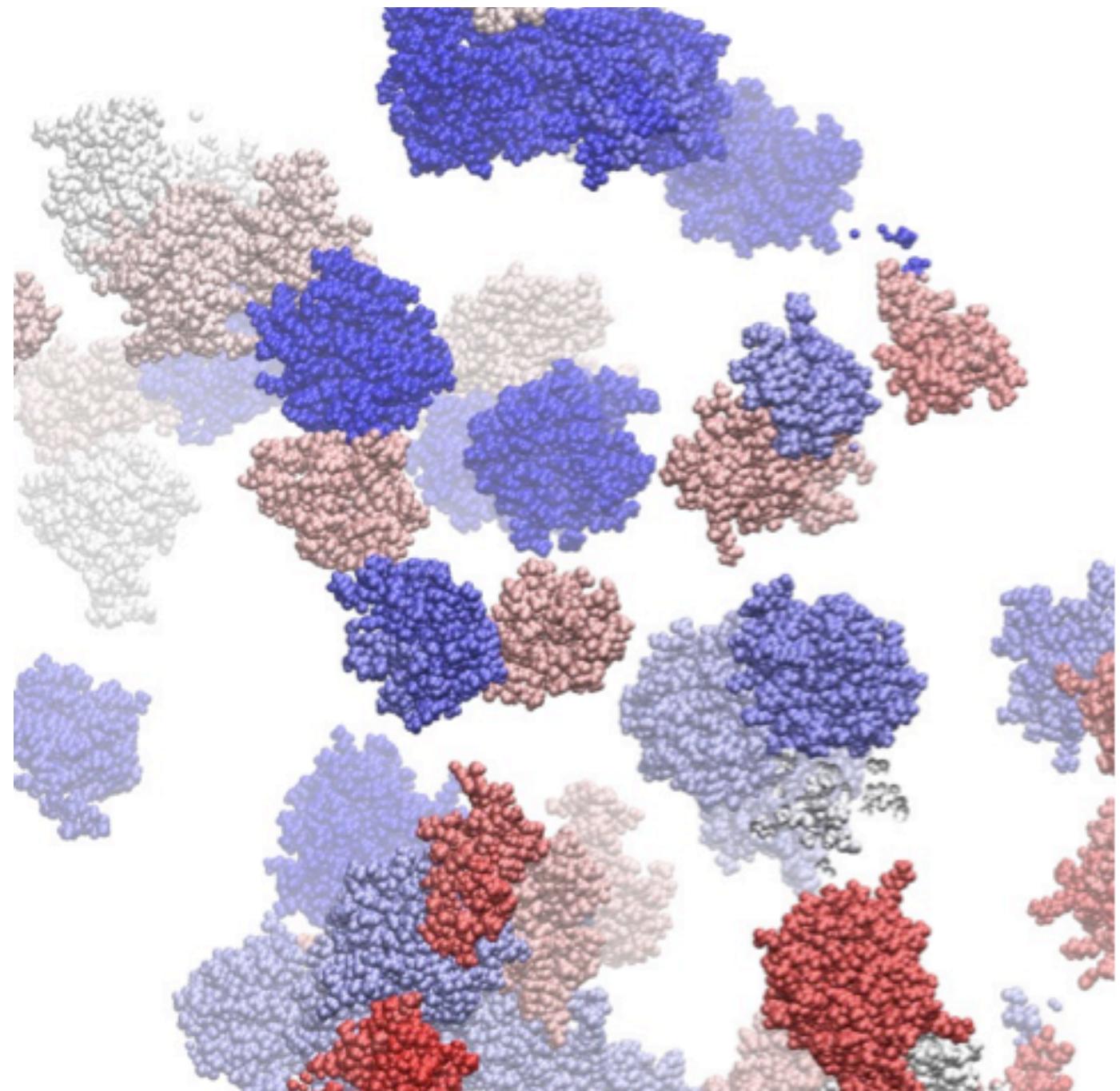
## **B) Protein-protein interactions (PPIs)**

- Resources (databases) and tools to study PPIs
- Protein-protein docking
  1. General principles (lecture)
  2. ClusPro
  3. Chimera
  4. ClusPro results analysis

## **C) Protein-small molecule interactions**

- Protein-small molecule docking
  1. General principles (lecture)
  2. Autodock VINA
  3. Virtual screening
- Drugs and drug targets
- Principles of drug design
- Principles of drug target prediction and drug repositioning
- Resources (databases) and tools to study protein-small molecule interactions

Molecular dynamics  
simulation snapshot of a  
colicin E9 and Im9 protein  
solution (van der Waals  
presentation and individual  
colors for each protein)  
indicating multiple  
transient association events  
Copyright: © Technische  
Universität München,  
Physik-Department T38



# Protein-ligand interactions

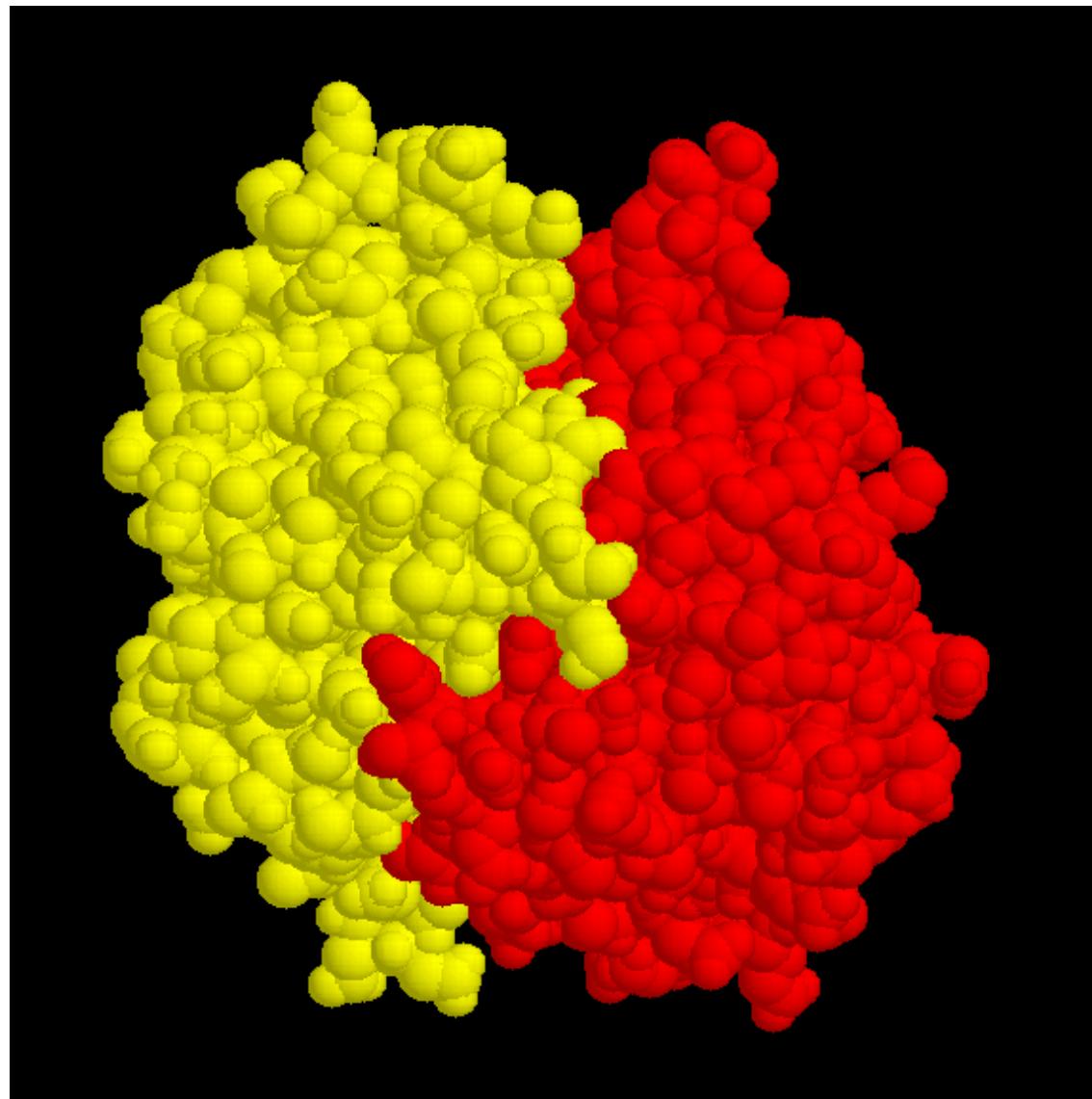
Protein functions are based on the capability of proteins to bind a certain ligand (or more)

- elemental ions
- small organic molecules (~600 Da or less)
- peptides
- macromolecules

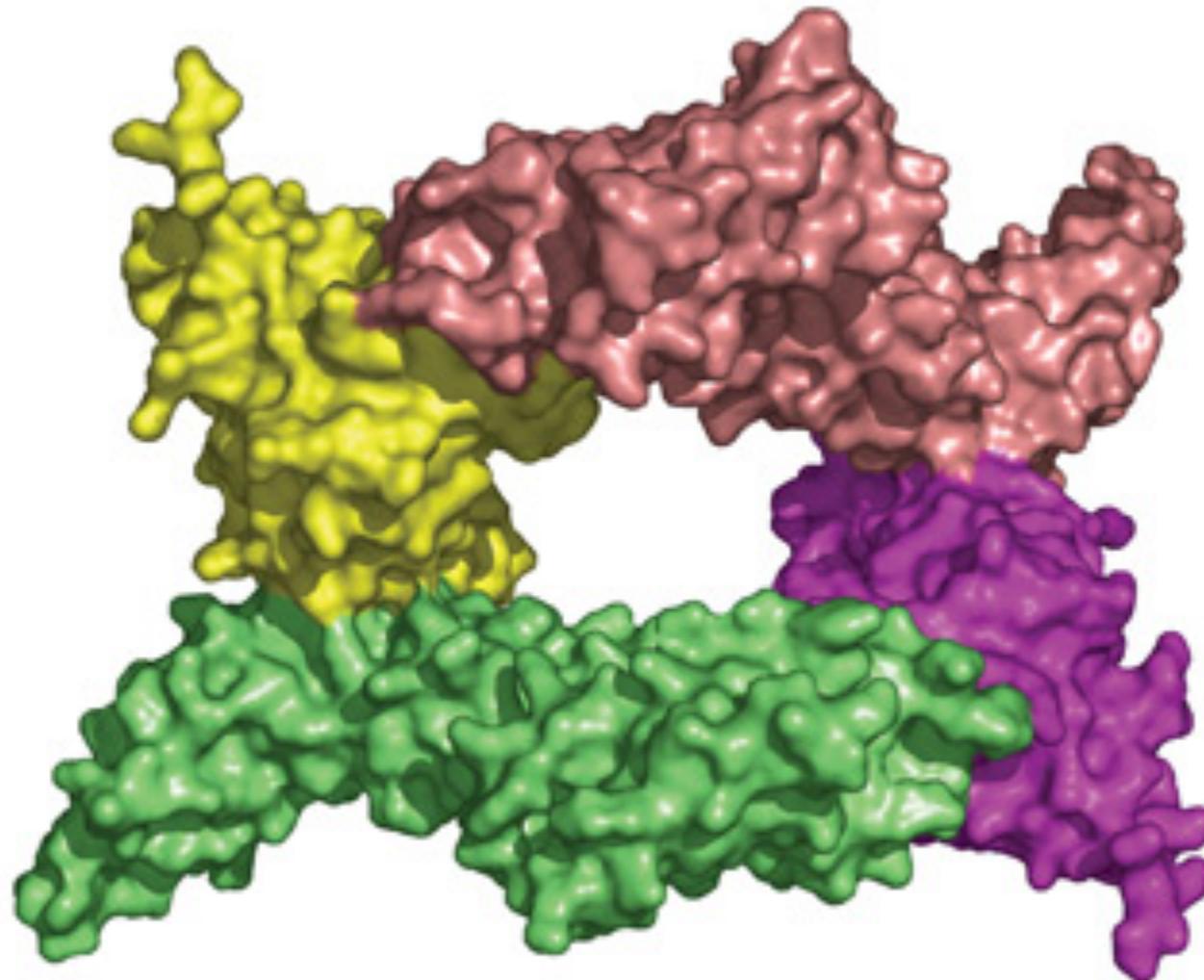
# A few definitions

- Binary interaction
- Protein complex
- Hub
- Homo vs hetero
- Stable vs transient
- Protein binding interface

# Binary interactions

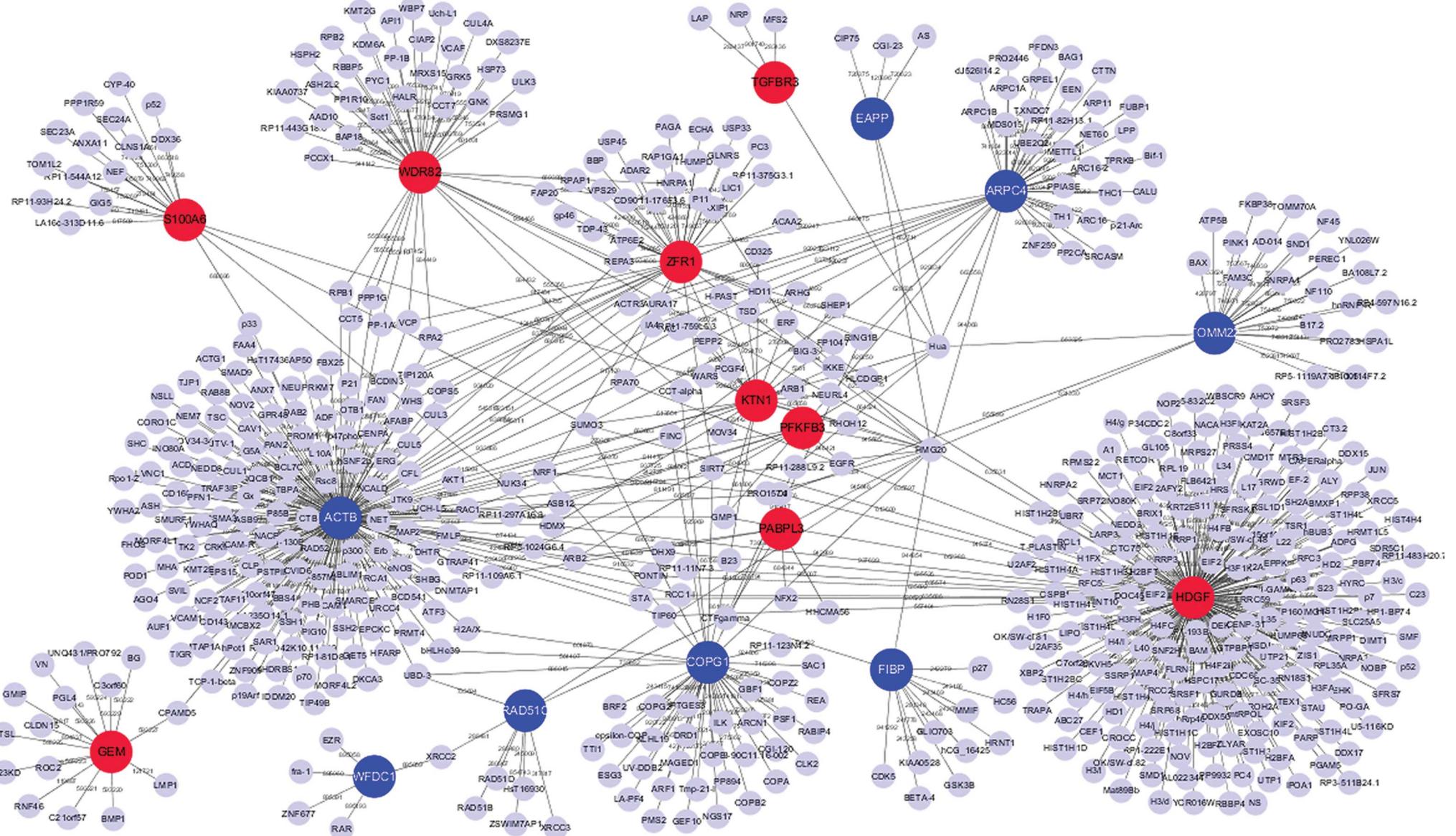


## Protein complex



A structural image of the protein-protein complex formed when YpkA (green and pink) binds to the Rac1 protein of the host cell (yellow and purple)

# Protein interaction network - hubs



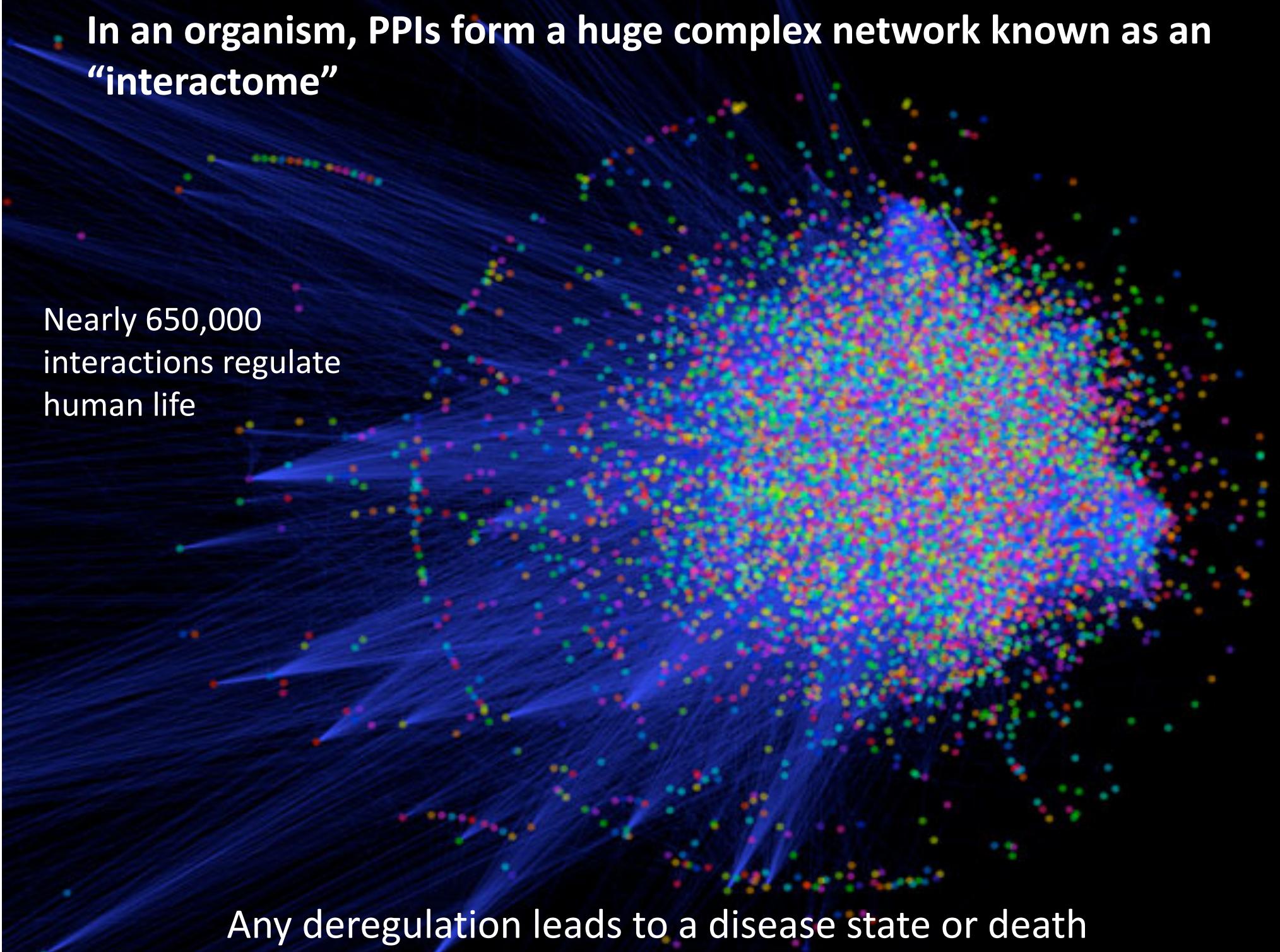
# PPIs are at the basis of key cellular processes

- Signal transduction/communication
- Enzyme-mediated catalysis
- Immune response
- Cellular division
- Programmed cell death
- Cell-cell recognition
- Viral action
- Metabolic and genetic regulatory networks

# Roles of ligands

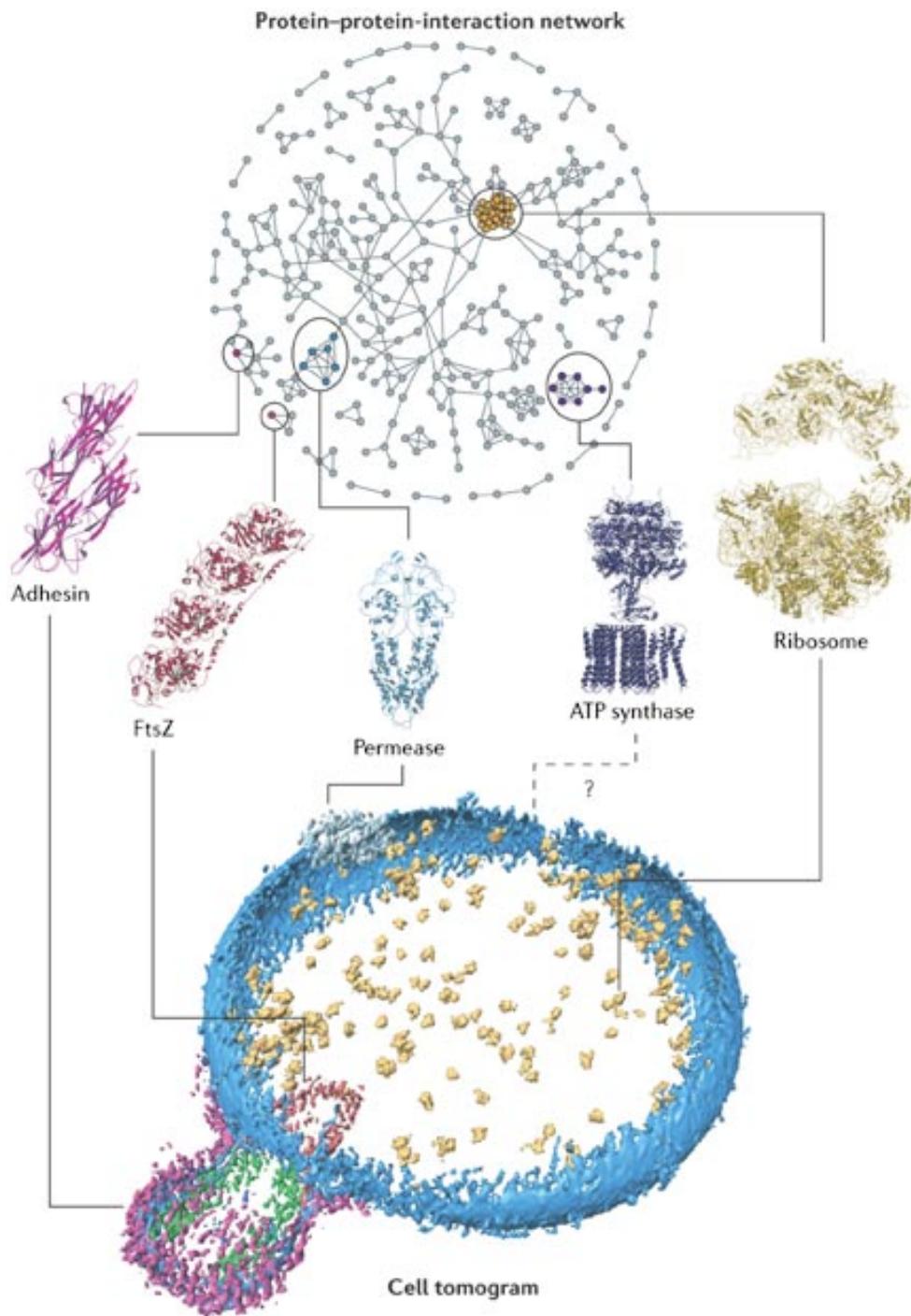
- Catalysis: substrate, products, cofactors
- Regulation: ATP, hormones, etc.
- Communication pathways: first or second messengers, downstream regulators
- Protein trafficking: ligands serving as means by which organelles or other macromolecules are recognised by proteins
- Prosthetic groups: heme group
- Defense/offense: toxins

- In an organism, PPIs form a huge complex network known as an “interactome”



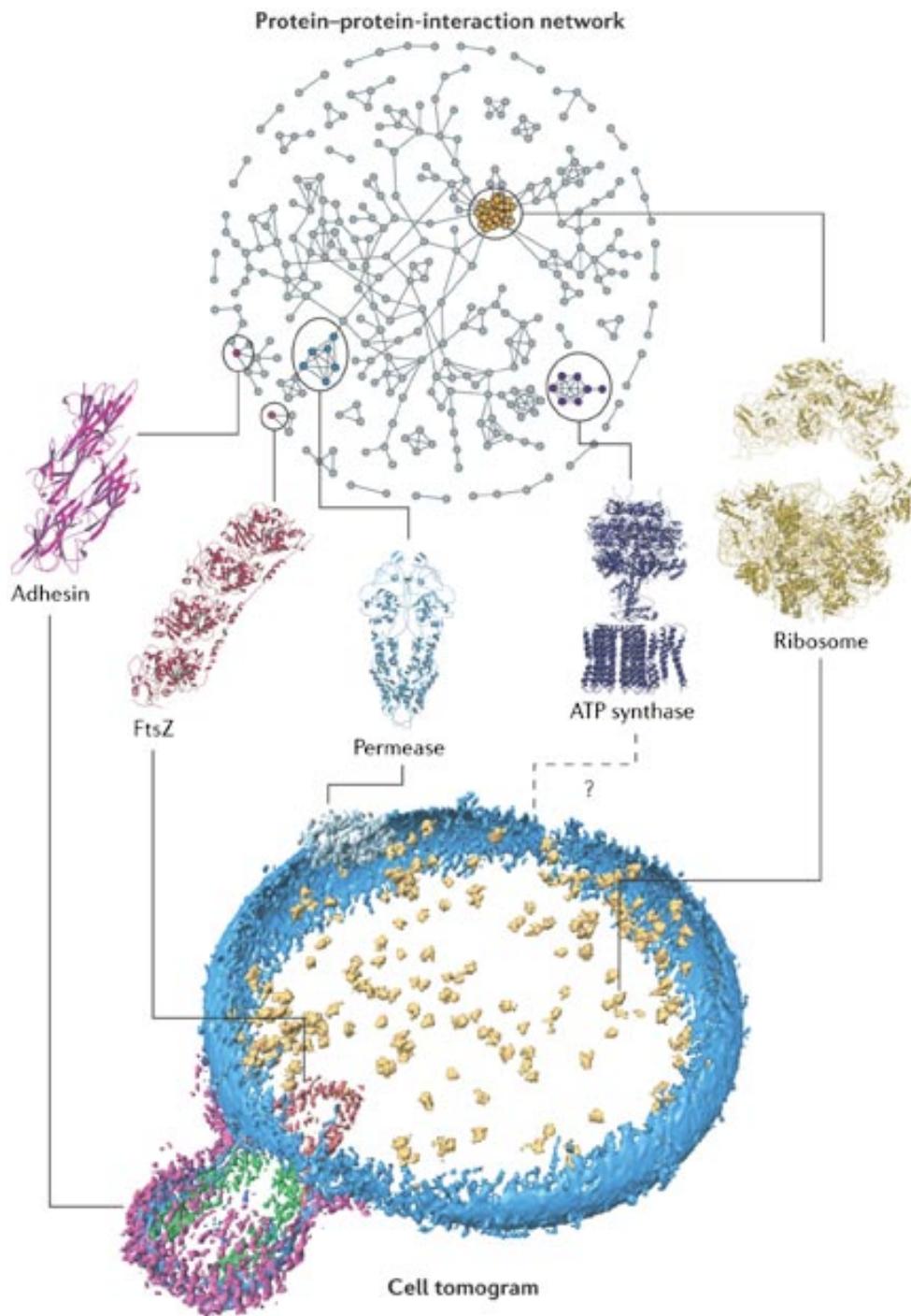
Nearly 650,000 interactions regulate human life

Any deregulation leads to a disease state or death



### Reductionist biology – molecular viewpoint

- **Specific molecule(s)** of interest
- Experiments to determine interaction partners and modes of interaction
- Prediction of interaction partners and modes of interaction
- Analysis of specific interaction

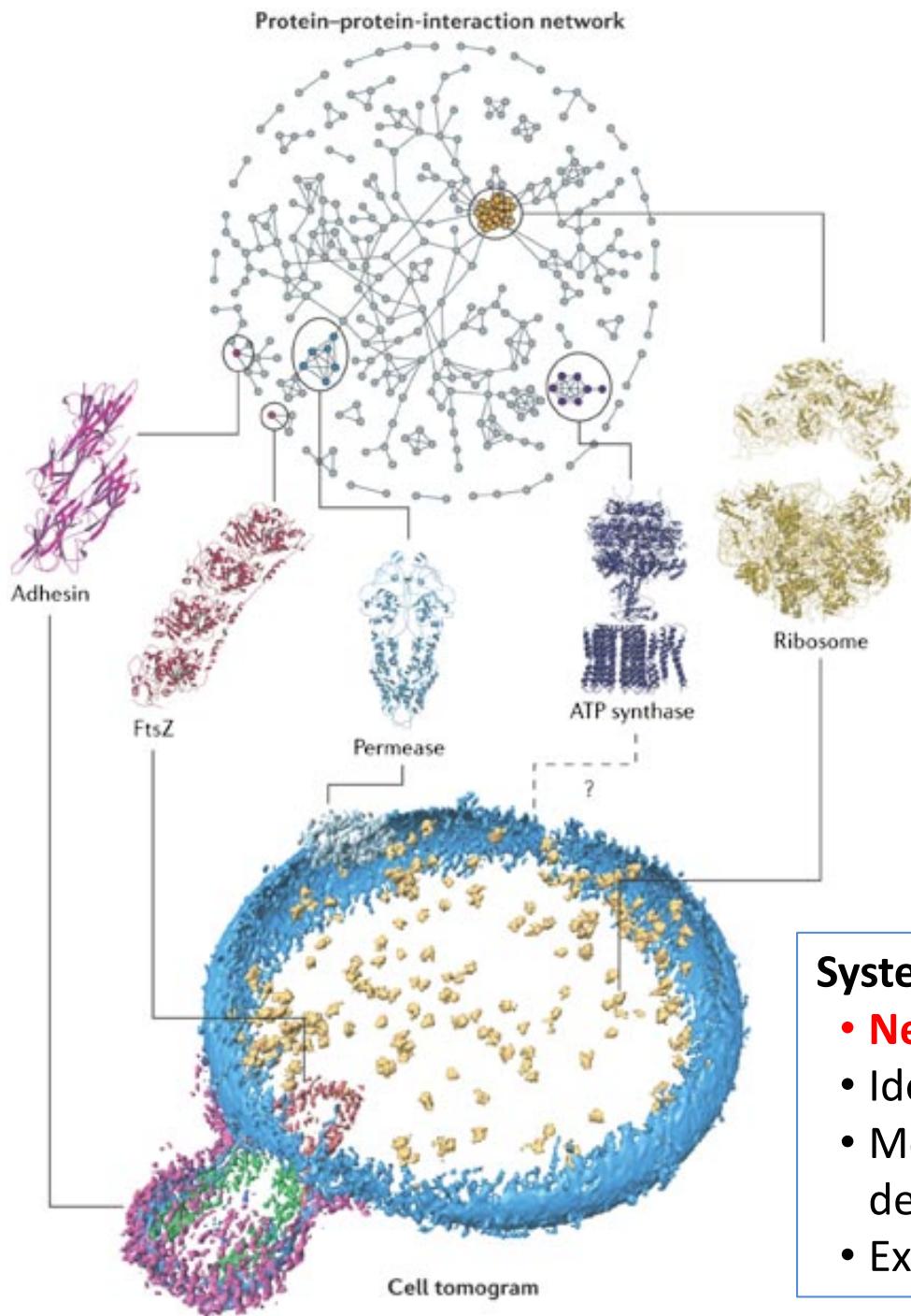


## Protein networks

- Identification of **functional modules** (set of proteins highly connected to each other)
- Hubs, singletons
- Networks' analysis -> biological hypotheses
- Prediction of new interactions

## Reductionist biology – molecular viewpoint

- **Specific molecule(s)** of interest
- Experiments to determine interaction partners and modes of interaction
- Prediction of interaction partners and modes of interaction
- Analysis of specific interaction



## Protein networks

- Identification of **functional modules** (set of proteins highly connected to each other)
- Hubs, singletons
- Networks' analysis -> biological hypotheses
- Prediction of new interactions

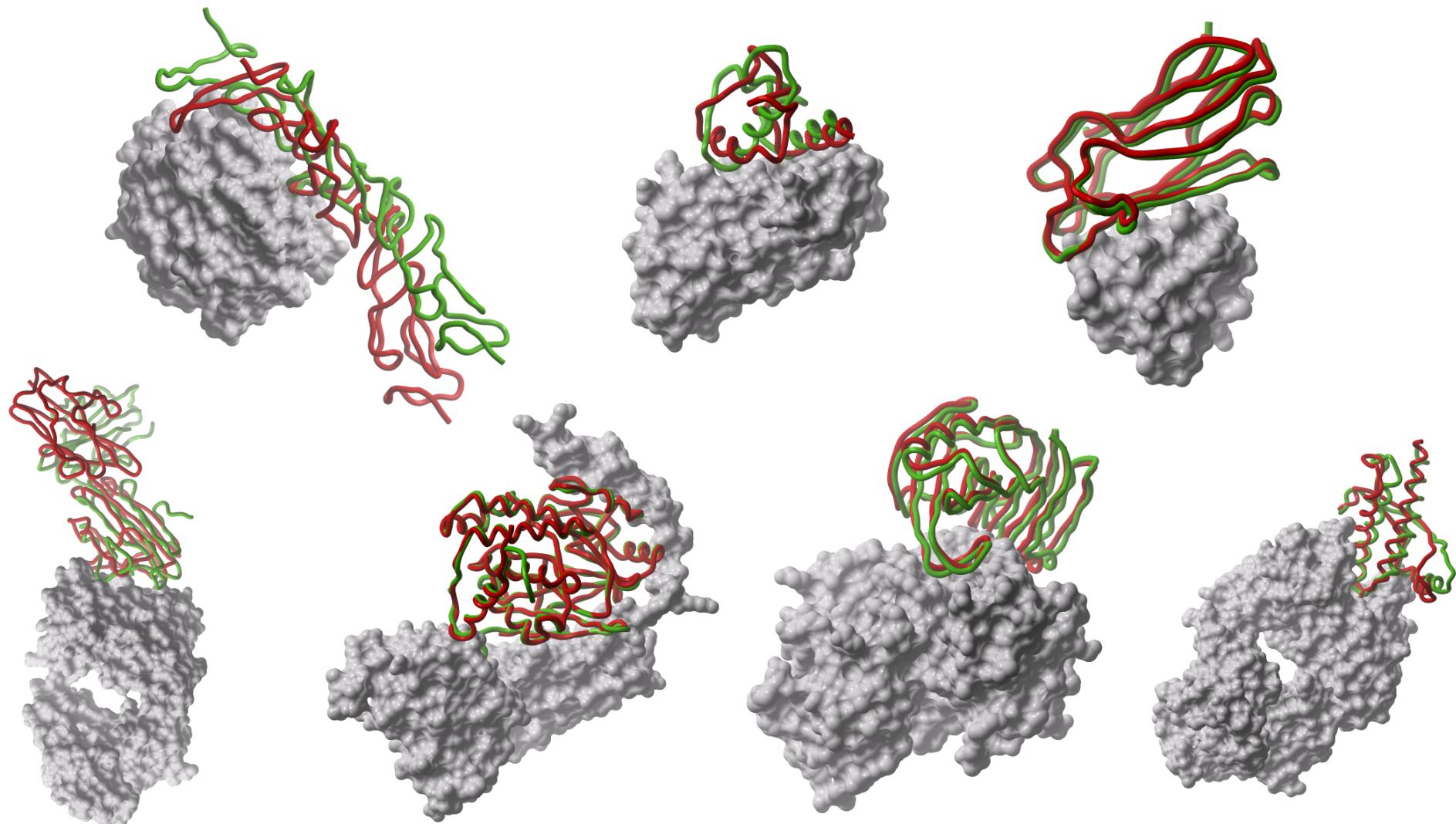
## Reductionist biology – molecular viewpoint

- **Specific molecule(s)** of interest
- Experiments to determine interaction partners and modes of interaction
- Prediction of interaction partners and modes of interaction
- Analysis of specific interaction

## Systems biology

- **Networks, pathways** implicated in a condition
- Identify perturbed or deregulated systems
- Modelling of the system to infer to signals and/or deregulation events
- Experiments to determine responses of the system

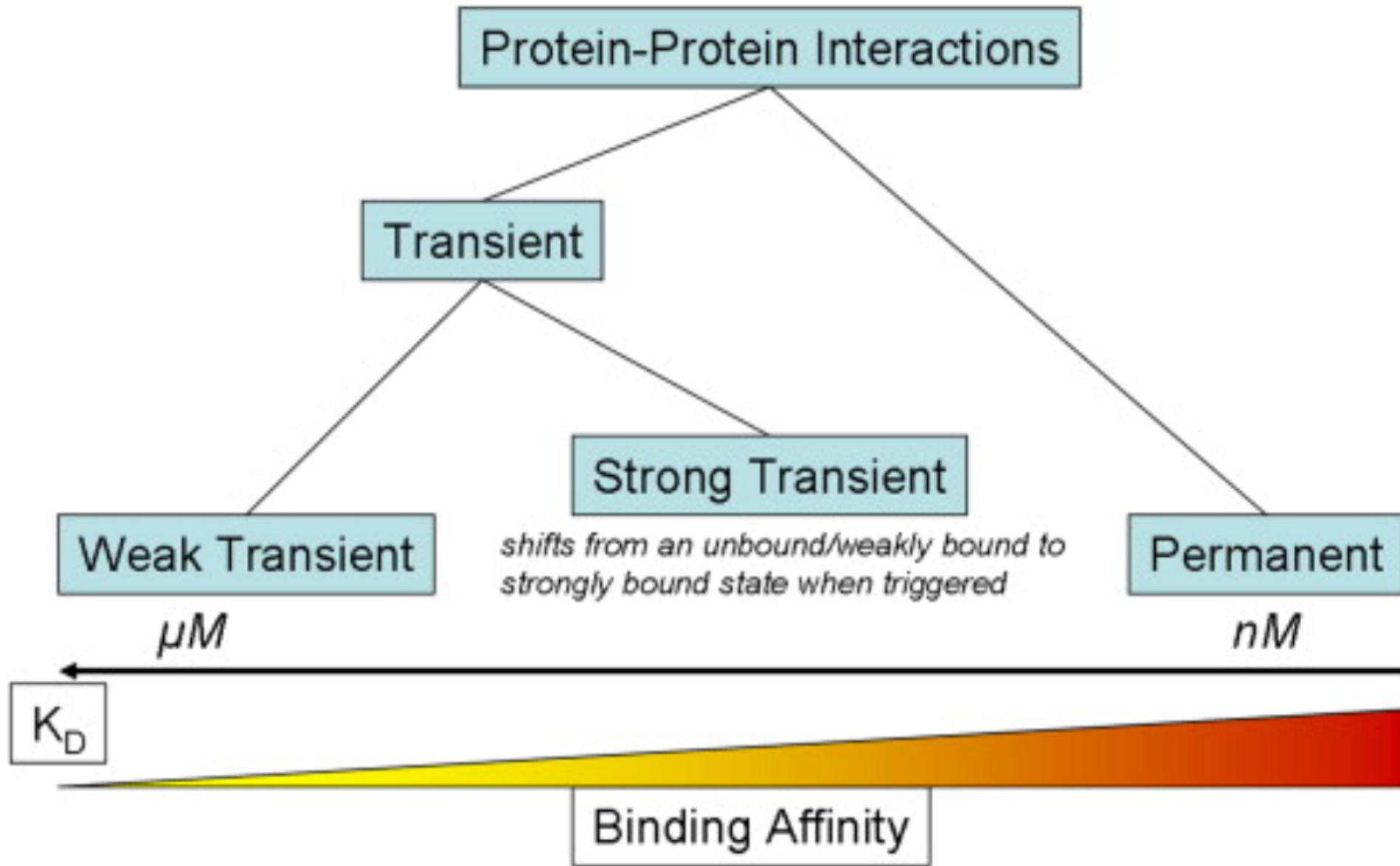
# The reductionist approach



# We can classify PPIs according to:

- Biological context: enzyme-substrate, antibody-antigen, receptor-hormone/neurotransmitter
- Permanence of binding: permanent-obligate vs transient
- Similarity of the binding partners: homo vs hetero
- Number of binding partners: hub, non-hub

Some of the above categories may partially overlap



**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

# Transient vs Stable

- Transient (relatively weak)
  - Brief and reversible interactions occurring in specific cellular contexts
- Stable (for a longer period of time)
  - Proteins take part of permanent complexes as subunits, in order to carry out structural or functional roles
  - Vocabulary: permanent, obligate, oligomeric, tight, more stable

# Examples

- Transient interactions
  - Interactions mediated by short linear motifs
  - Interactions mediated by PTMs
  - Disorder-to-order transitions
  - Proteins involved in signalling cascades
- Stable interactions
  - Homo-oligomeric or hetero-oligomeric complexes
  - Interactions mediated by PTMs
  - enzyme-inhibitor
  - antibody-antigen
  - domain-domain
  - domain-peptide

# We can classify PPIs according to:

- Biological context: enzyme-substrate, antibody-antigen, receptor-hormone/neurotransmitter
- Permanence of binding: permanent-obligate vs transient
- Similarity of the binding partners: homo vs hetero
- Number of binding partners: hub, non-hub

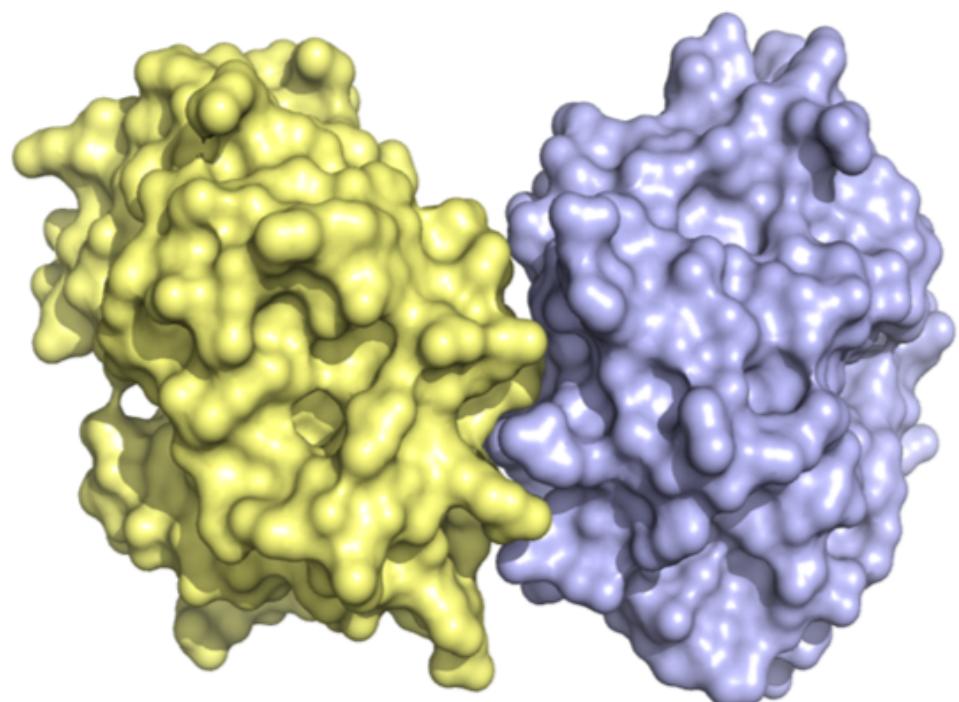
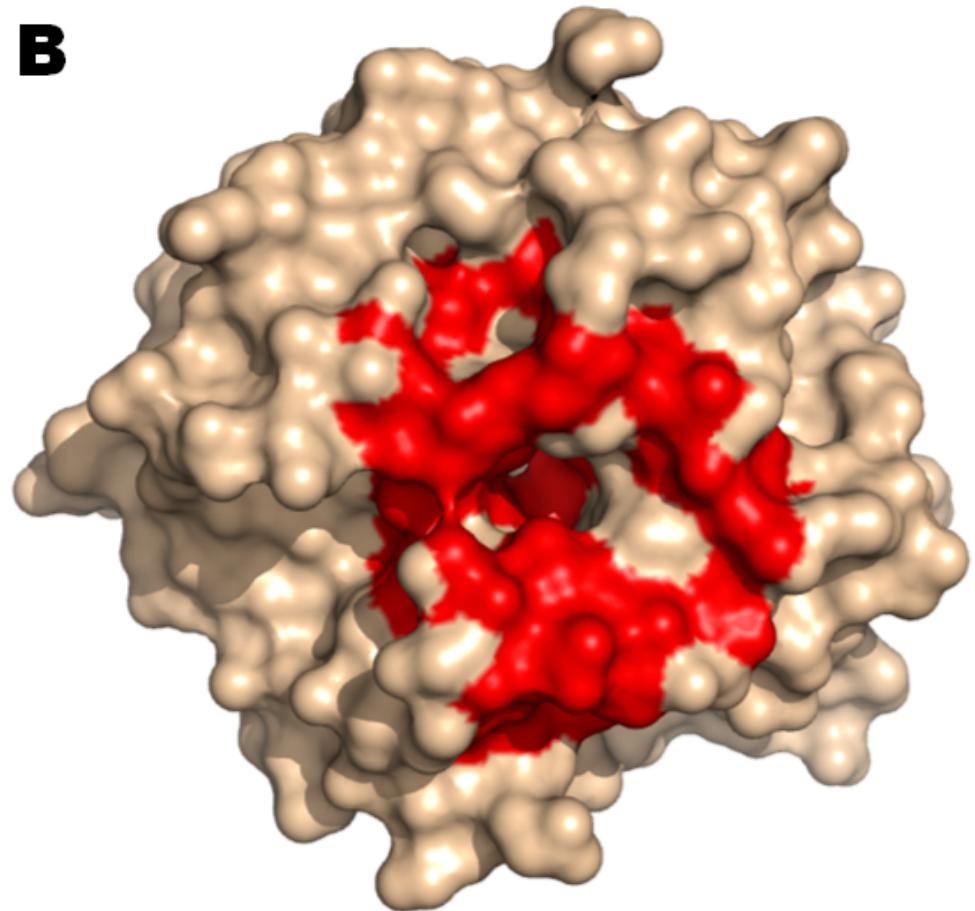
Some of the above categories may partially overlap

# Protein-ligand interfaces

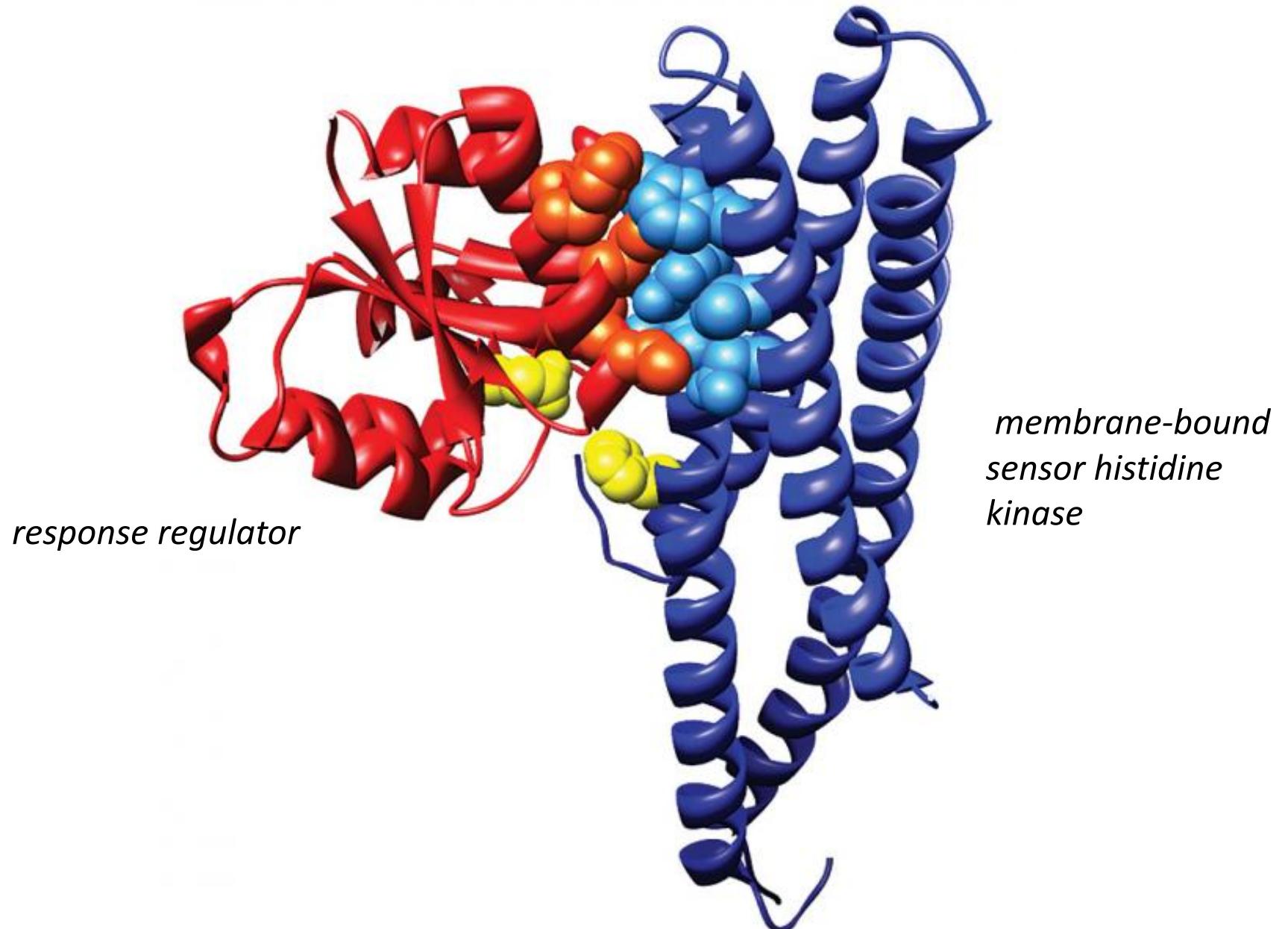
## Structure-function relationship

The ability of proteins to form **biologically active complexes** depends on the properties of their binding surfaces:

- Size
- Geometric compatibility
- Chemical composition
- Polarity
- Atom packing efficiency
- Hydrogen bond or salt bridge frequency
- Number of buried water molecules
- Interaction energy
- Residue conservation
- Types of secondary structures

**A****B**

**A.** Hpx14 homodimer (PDB 3C7X) **B.** Definition of the docking site (red patch)



. The catalytic residues that exchange a phosphoryl group are shown in yellow (courtesy of Alexander Schug and Hendrik Szurmant)

What is the protein-ligand interaction  
driving force?

Biological systems are not isolated: they exist in a state of constant **pressure** and **temperature**

The thermodynamic quantity used to determine the direction of a process is the Gibbs free energy G

The direction of a process is determined by measuring the variation of G:

$$\Delta G$$

$\Delta G = 0$  at the equilibrium

$\Delta G > 0$  endergonic processes

$\Delta G < 0$  exergonic processes



Spontaneous processes proceed decreasing their free energy

The variation of free energy can be written in terms of enthalpy (H) and entropy (S) changes at a **constant temperature**

$$\Delta G = \Delta H - T\Delta S$$

Enthalpy: amount of energy that a system can exchange with the environment

$\Delta H > 0$  mainly corresponds to the rupture of favourable non-covalent bonds

$$H = E + PV \Rightarrow \Delta H \approx \Delta E$$

- A. Formation or breaking of covalent bonds
- B. Variations in the electrostatic or van der Waals interactions
- C. Thermally induced changes in the atomic motions

$\Delta H$  = Change in enthalpy. It represents the general tendency to minimise the energy

$\Delta S$  = Entropy change. It represents the general tendency to maximise disorder

# Entropy S

---

$$S = k_B \ln \omega$$

$\omega$  = number of the possible configurations of the system

The number of possible states is inversely related to the order in the system. A highly ordered system can only access a few states while a disordered system can sample many

In the binding to a protein, the atoms of a ligand gradually become more confined and the **entropy decreases**.

However, the binding can occur thanks to the **hydrophobic effect**, which is accompanied by **an increase in entropy of the solvent**.

# What happens in protein folding?

In the process of folding of a polypeptide chain the increase in entropy of the solvent compensates the loss of entropy of the macromolecule, leading to a general increase in entropy of the system as a whole.

$$\Delta G_{\text{folding}} = \Delta H - T\Delta S$$

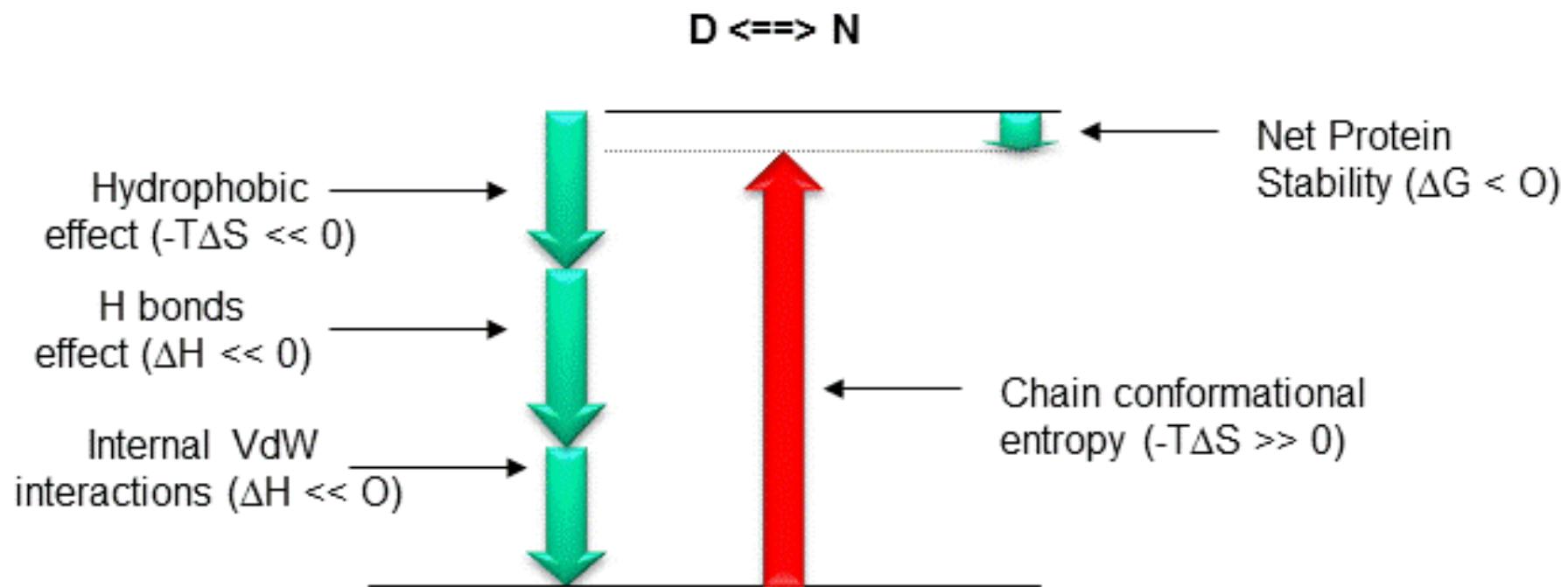
$$\Delta H < 0$$

(formation of the  
electrostatic or van  
der Waals  
interactions )

$$T\Delta S_{\text{conformational}} < 0$$

$$T\Delta S_{\text{hydrophobic effect}} > 0$$

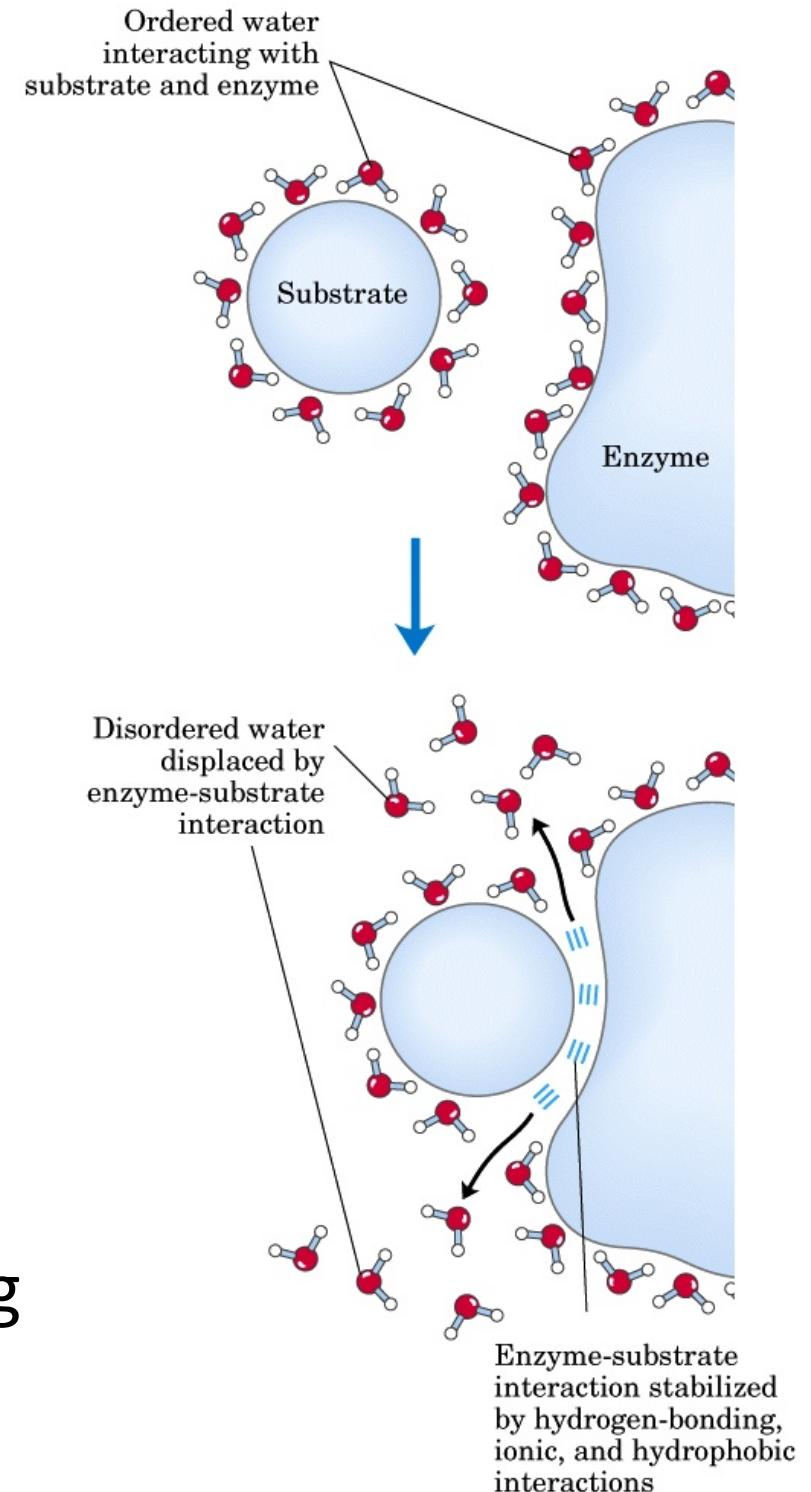
## Thermodynamics of Protein Folding



The driving force behind many events within the cell is represented by the increase in **entropy generated in the water molecules when a reaction occurs**

$$\Delta G = \Delta H - T\Delta S$$

- Ligand binding is driven by the **hydrophobic effect** and **van der Waals interactions**, whereas electrostatic interactions make them specific
- PPI interfaces tend to be more **hydrophobic** than surfaces that do not participate in the binding
- Studies on enzymes demonstrate the importance of the hydrophobic effect also in binding non-protein substrates



# Forces governing biomolecular recognition

- Van der Waals
- Electrostatics
- Hydrophobic contacts
- Hydrogen bonds
- Salt bridges

All interactions act at short range → surface complementarity is needed for tight binding

# The Binding Site (BS)

- Geometry alone is insufficient to recognise the correct ligand
- The binding site-ligand match is based on:  
**geometry** and **electrostatics**
- Geometric and electrostatic complementarity can be achieved using different BS architectures

# Geometric complementarity

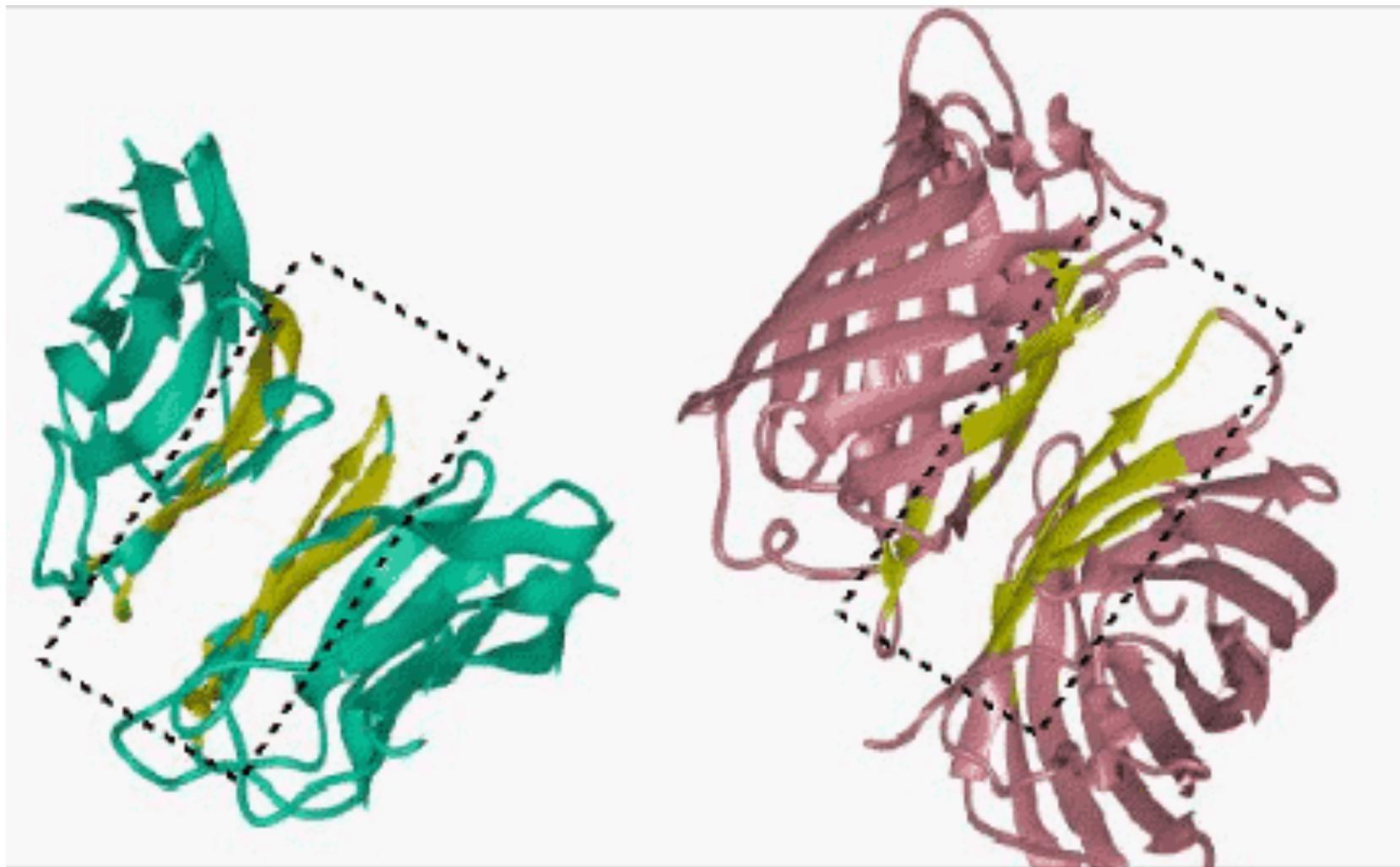
- The geometrical match optimises all non-covalent interactions that mediate the binding
- Short range and Van der Waals

# Electrostatic complementarity

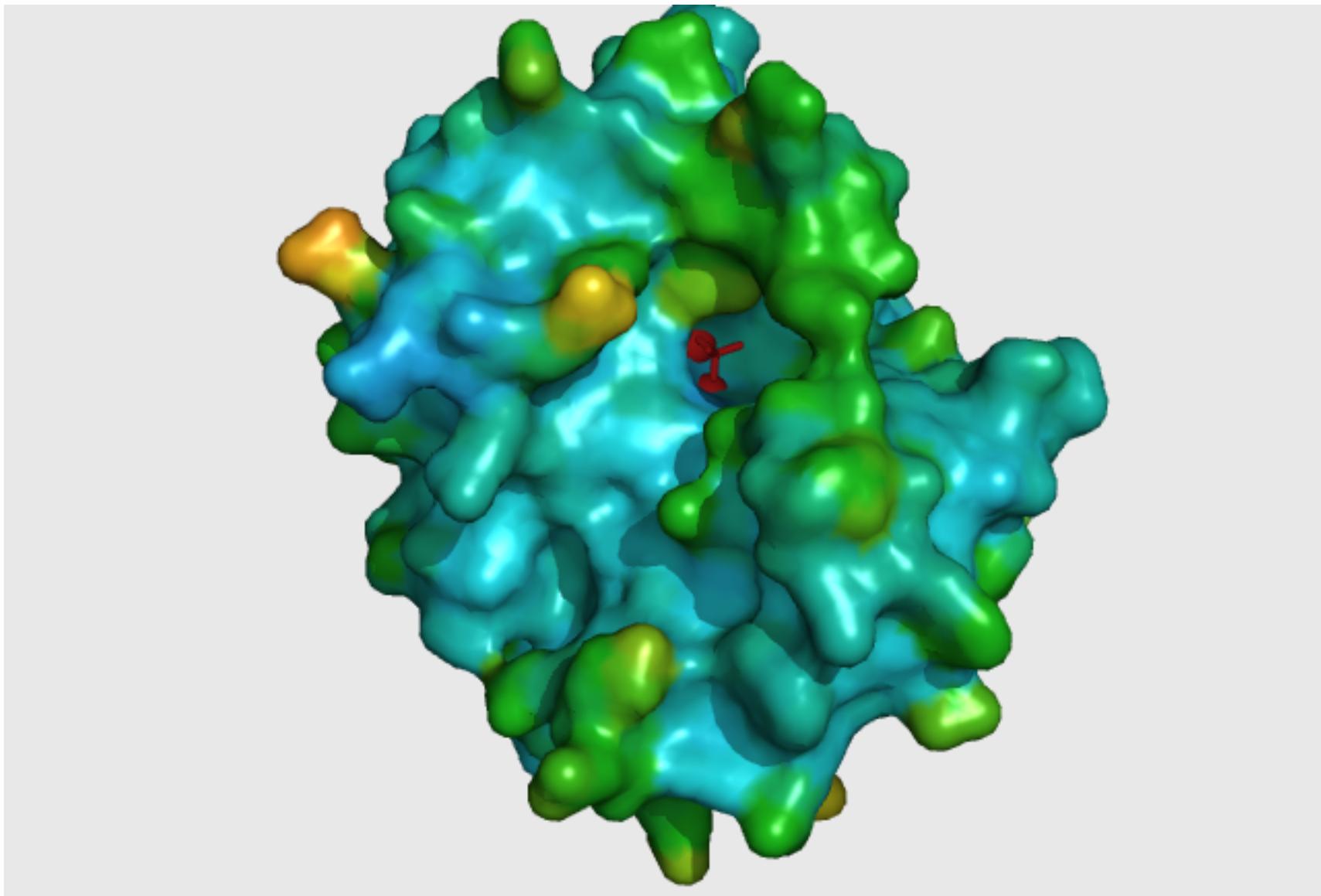
- When ligand carries an electric charge → The BS tends to contain the opposite charge
- Electrostatic match increases binding **specificity**

# PPI surface size

- The area of PPI interfaces is large (1000 to 4000 Å<sup>2</sup>)
- **Standard-sized** interfaces are 1200 to 2000 Å<sup>2</sup>
- **Short-lived and low-stability complexes** -> smaller interfaces (1150–1200 Å<sup>2</sup>)
- **Protein-small molecule interaction** surfaces have a smaller area (300 - 1000 Å<sup>2</sup>).



# Protein-small molecule: small and deep depressions



# Other PPI surface features

- **Flatter** than that of other protein-ligand interfaces
- Lack the grooves and pockets observed at the surfaces binding small molecules
- Generally **hydrophobic**
- Tighter atom packing
- Lower polarity
- Higher conservation (especially at the center of the interface)

# Some more facts about PPIs

- Transient interfaces differ from stable interfaces
- Stable interfaces: less planar, more hydrophobic, better geometric and electrostatic complementarity
- Homodimeric interfaces resemble protein cores
- Heterodimeric interfaces are more like regular (i.e. non-binding) surfaces

# Specificity

- Conferred by electrostatic interactions
- Electrostatic interactions prevent aggregation  
(mediated by non-polar interactions)

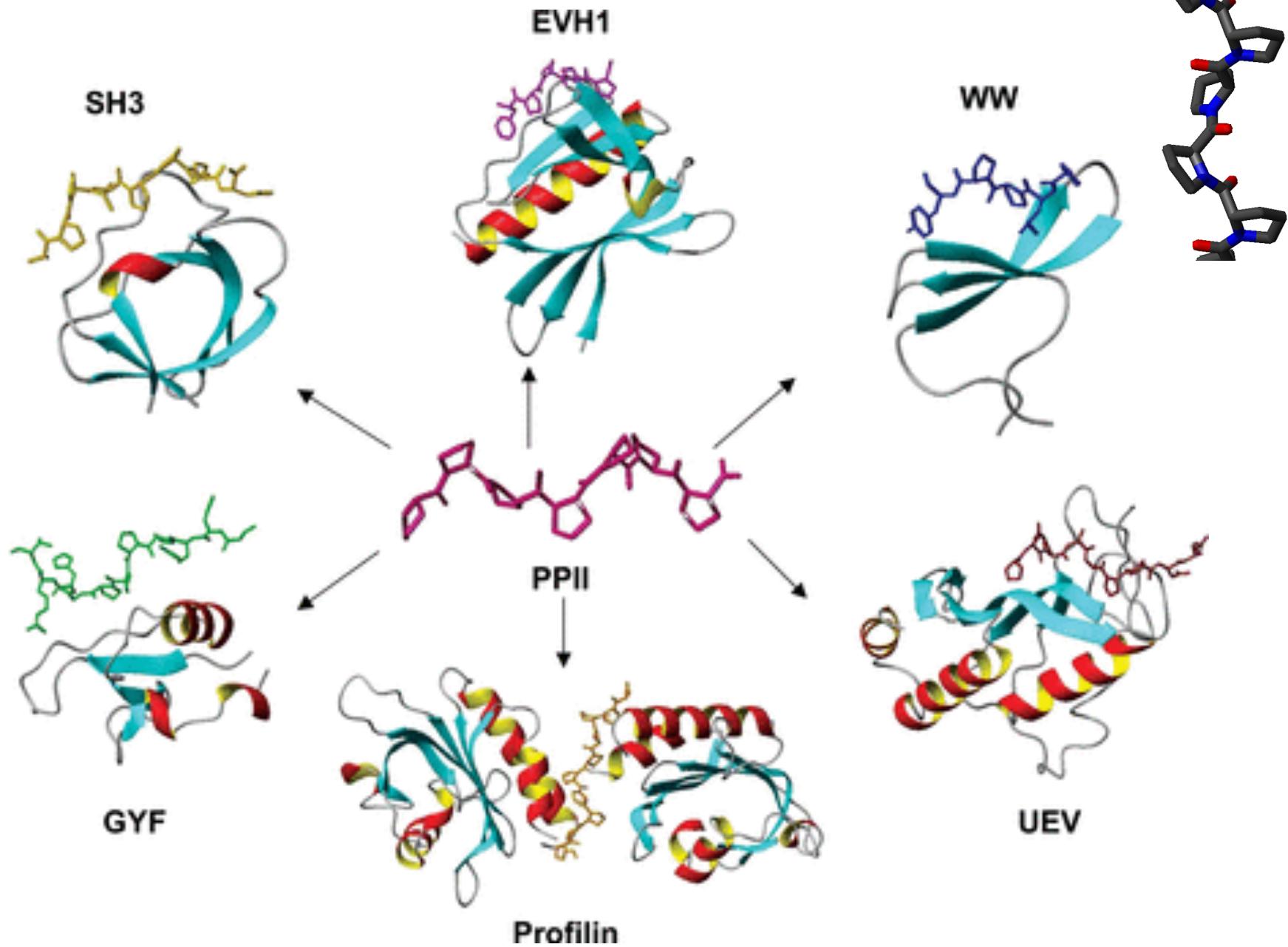
# Affinity

- 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 can maintain a high degree of specificity for their partners

# Specificity

- Fixed scaffold, change of loop structure (antibodies)
- Many proteins exhibit affinity for multiple partners (Reichmann et al. Curr. Opin. Struct. Biol. 2007) → hubs
- The ability to bind different partners → Low specificity
- Different proteins can bind the same interface, each interacting with a different cluster of **nonpolar** residues OR
- To the same **highly flexible** interface

# Polyproline type II helices



# Protein-ligand binding energetics

## Protein-ligand affinities

- Binding energies may range from -2.5 to -22 kcal/mol
- Signal transduction networks -> weak interactions
- Cofactor binding by enzymes (reversible but strong) -> -5.5 to -9.5 kcal/mol
- Antibody-antigen:  $-8 \pm 3$  kcal/mol
- Receptor-hormone: -12 kcal/mol
- Enzyme-inhibitor:  $-12 \pm 3$  kcal/mol
- Enzyme-transition state:  $-22 \pm 5$  kcal/mol

# Hot spots

- The protein-protein binding interface is heterogeneous
- Only certain hydrophobic spots contribute to the free energy of binding and help to hold the two proteins together
- Affinity seems to result from a small number of residues
- Such residues are called **hot spots**

**Hot spots** are defined as residues whose replacement by alanine leads to an affinity/free energy change of at least 2 kcal/mol

# Some fact about **hot spots**

- Hot spots account for less than 50% of the contact area of PPI
- Hot spots tend to be **evolutionary conserved**
- Hot spots tend to appear as clusters in the interface
- Some residues are found more frequently in hot spots (e.g., Phe, Tyr, Trp)
- Hot spots are surrounded by energetically less important residues that probably separate/prevent bulk water from hot spots

# Molecular structure of stable interactions (complexes): what information?

- 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

# How PPI surfaces can be studied?

## Experimental:

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

## Computational:

- Prediction of interaction/binding sites
- Prediction of protein-protein complexes
- *In silico* mutational studies

# Computational alanine-scanning in protein-protein interfaces

**DrugScore PPI**

*In silico* alanine scanning

Email  optional (see notes below)

PDB-ID  OR  [Sfoglia...](#) (\*)

Chain(s)  (comma separated)

 [Can't read?](#)

Via submit, you have read and accepted the [terms and conditions](#).

© Dennis M. Krüger, 2009

 ROBETTA BETA  
Full-chain Protein Structure Prediction Server

---

Structure Prediction Fragment Libraries Alanine Scanning DNA Interface Scan  
[ Queue ] [ Submit ]  
[ Register / Update ] [ Docs / FAQs ] [ Login ]

**Submit a job to the Computational Interface Alanine Scanning Server**

Please do not submit more than 10 targets at a time

**Required**

[Registered Username:](#) or [Registered Email Address:](#)

Job Name:

**Warning! Complex will be publicly accessible.**

Upload Complex:  [Sfoglia...](#)

<http://robetta.bakerlab.org/alascansubmit.jsp>

<http://cpclab.uni-duesseldorf.de/dsppi/>

**DrugScorePPI** and **Robetta Alascanning** are dedicated to the identification of hot spots at protein-protein interfaces

<http://prism.ccbb.ku.edu.tr/hotpoint/>

The screenshot shows the HotPOINT web interface. At the top left is the logo 'HotPOINT HOT SPOT PREDICTION SERVER FOR PROTEIN INTERFACES' with the KOC UNIVERSITY logo. A navigation menu on the left includes links for Home, Documents, Links, and About. Below the menu is a citation: 'Tuncbag N, Gursoy A, Keskin O. Identification of computational hotspots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy. *Bioinformatics*, 2009 Jun 15;25(12):1513-20. [\[Link\]](#)'. A note below says: 'Below you can try our prediction algorithm by entering the four letter PDB code of a protein or uploading your own structure file that is in the PDB format with the chain identifiers. Please do not submit PDB files containing only one chain. This will return an error! Hotpoint requires two chain identifiers which corresponds to a protein interface.' There are three input fields: 'Run our prediction algorithm for a particular input protein.', 'Enter the four letter PDB code for automatic download from PDB: ', and 'Or load your structure file from disk:  Stopfile'. Below these is a Jmol visualization window titled 'Interactive Jmol visualization for 1AXDAB'. The visualization shows two protein chains, A and B, represented by blue and yellow sticks. The interface residues for chain A are highlighted in red, and the interface residues for chain B are highlighted in orange. Text boxes at the top of the visualization area say 'Interface residues on chain A' and 'Predicted Hot spots in interface (Chain A)'. Another text box on the right says 'Interface residues on chain B' and 'Predicted Hot spots in interface (Chain B)'.

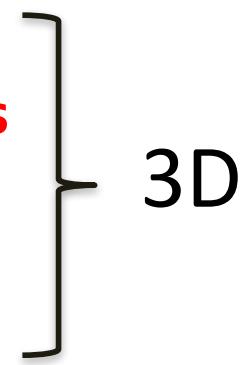
**HotPOINT** calculates solvent accessibilities and total contact potentials. It predicts hot spots at the complex interface

**PIC** recognises various kinds of interactions, such as disulphide bonds, hydrophobic interactions, ionic interactions, hydrogen bonds, aromatic- aromatic interactions, aromatic-sulphur interactions and cation- $\pi$  interactions between proteins in a complex

<http://crick.mbu.iisc.ernet.in/~PIC/>

The screenshot shows the PIC web interface. At the top center is the title 'P I C : Protein Interactions Calculator' and the subtitle 'Molecular Biophysics Unit, Indian Institute of Science, Bangalore.' Below this is a horizontal menu bar with links: HOME, HELP, CRITERIA, SUBMIT JOB, CONTACT US, and LAB PAGE. A note below the menu says: 'You can also submit your jobs on the [new link](#).' A section titled 'INTRODUCTION:' contains the following text: 'Protein Interactions Calculator (PIC) is a server which recognizes various kinds of interactions; such as disulphide bonds, hydrophobic interactions, ionic interactions, hydrogen bonds, aromatic- aromatic interactions, aromatic-sulphur interactions and cation -  $\pi$  interactions within a protein or between proteins in a complex. It also determines the accessible surface area as well as the distance of a residue from the surface of the protein. The input should be in the Protein data bank(.pdb) format. Interactions are calculated based on empirical or semi-empirical set of rules.'

# Prediction of protein-protein interactions

- Phylogenetic profiling
  - Prediction of co-evolved protein pairs based on similar phylogenetic trees
  - Rosetta stone method
  - Association methods
  - Bayesian network modelling
  - Domain-pair exclusion analysis
  - Supervised learning problem
  - Gene fusion
  - **Classification methods**
  - **Inference of interactions from homologous structures**
  - **Identification of structural patterns (hot spots)**
  - **Protein-protein docking**
- 
- 3D



# **Energy and work are necessary to keep order**

---

- The entropy of the universe tends to grow
- The formation of macromolecules is not an accidental process but requires work
- The synthesis of macromolecules needs energy

We measure the energy variations occurring during chemical reactions in terms of GIBBS FREE ENERGY, G.

# **Work and energy are necessary to keep order**

---

G represents the “usable” energy of a system, i.e. the energy available to carry out work

$$\mathbf{G = H - TS}$$

In particular, G is the thermodynamic quantity that can be used to determine the direction of a process

## How?

---

Gibbs free energy is a state variable: it depends on the initial and final equilibrium states and not on the way in which the system acquired that states.

To decipher the direction of a process, we can look at its initial and final state, i.e. evaluate the free energy variation:

$$\Delta G$$

The free energy variation can be written in terms of enthalpy (H) and entropy (S) change at constant temperature.

$$\Delta G = \Delta H - T\Delta S$$

$\Delta H$  = Enthalpy variation. It represents the heat released or absorbed by the system in a reaction.

$\Delta S$  = Entropy variation. It represents the general tendency of a system to maximise its disorder.

# Enthalpy H

---

Quantity of energy that a system can exchange with the environment

$$H = E + PV$$

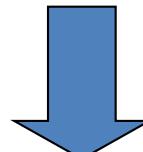
Internal energy  
of the system

pressure

volume

Biological systems are either solid or liquid. Therefore they do not usually undergo volume or pressure changes.

Biological systems are not isolated: they exist in a state of **constant temperature and pressure**.



$$\Delta H \approx \Delta E$$

# Internal energy E

---

$$E = U + K$$

Potential energy

Kinetic energy

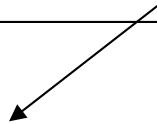
Position energy: in a macromolecule the potential energy is the result of all field (electrostatic) effects of the atoms of the molecular system

It is the result of the molecular motions induced by the heat



Set of covalent and non-covalent interactions in a molecular system.

$$\Delta G = \Delta H - T\Delta S$$



- A. covalent bond breaking or making
- B. Variation in the electrostatic or van der Waals interactions
- C. Variations induced by the heat in the atomic motions

These variations at constant pressure entail heat transfer between the system and the environment. Therefore, we can write:

$$Q_{(p)} = \Delta H \approx \Delta E$$

Therefore, at constant pressure and volume, the enthalpy variation coincides with both the heat and the variation of internal energy occurred during the process.

**$Q_{(p)}$  can be experimentally measured**

Conventionally

$\Delta H > 0$  in a reaction **adsorbing** heat  
 $\Delta H < 0$  in a reaction **releasing** heat

Energy is **released** to make bonds. The enthalpy change is **negative** because the system is **releasing** energy when forming bond.

$$\Delta H < 0$$

Energy is **required** to break bonds. The enthalpy change for breaking bonds is **positive**.

$$\Delta H > 0$$

# Entropy S

---

$$S = k_B \ln \omega$$

$\omega$  = number of possible configurations of the system (microstates)

The number of possible states is inversely proportional to the system order

In classic thermodynamics, at constant temperature, entropy and enthalpy fulfill the following relation:

$$\Delta S = \frac{\Delta H}{T} = \frac{Q_{(p)}}{T}$$

Changes of G determine the direction of a process:

$\Delta G = 0$  at the equilibrium

$\Delta G > 0$  **endergonic** processes (requiring energy)

$\Delta G < 0$  **exergonic** processes (releasing energy)

Spontaneous processes decrease their free energy

# Theories on protein-ligand binding

- **Lock and key theory** (enzyme-substrate)
- **Induced fit theory** (enzyme-substrate)
- **Monod-Wyman-Changeux** (MWC) model or pre-existing equilibrium theory (two different equilibrium conformations)
- **Population shift theory**: proteins shift spontaneously between multiple conformations (substates). The ligand stabilises preferentially one.