Laboratory of Bioinformatics 1 part B - Allegra Via

Saul Pierotti

March 15, 2020

Protein interaction and properties of binding sites

- The binding sites for small molecules tend to be small and deep
- Protein-protein interactions (PPIs) can be approached in from different perspectives
  - The reductionist approach focuses on specific molecules of interest and analyses specific interactions
    - It aimes at predicting parteners and modes of interaction
    - It requires wet-lab experiments
  - The protein network approach focuses on a set of proteins that interact with each other
    - It relies on the underlying biology and focuses on predicting new interactions
    - It can identify hubs of proteins that frequently interact with each other
    - A singleton is a protein that has only one interaction partner
  - The system's biology approach uses matematical models that rely on differential equations
    - The inputs to the equations are parameters that describe the biological system, such as pH and redox potential
    - It focuses on perturbation to the system
    - It can rely on experiment to determine the effect of a perturbation
- Interactions can be among proteins, with other non-protein macromolecules or with small molecules
  - In this course we will focus on interactions among proteins (PPIs)
- Several ways to classify PPIs
  - Interactions can be direct or indirect
  - Binary interactions or protein complexes (2 or many interactors)
- The protein binding interface is defined as the set of atoms of a protein which are less than 5 Å apart from an atom of the partner protein
- PPIs are an essential part of signal transduction, enzimatic activity, cellular division, metabolic networks
- Protein ligands can take part in catalitic mechanisms, regulation, and other biological activities
- PPIs can be classified depending on the biological context, binding permanence, similarity of binding partners and number of partners
- Transient interaction persist for a short time, while permanent interactions are more persistent
  - Transient interactions are typically mediated by short linear motives, PTMs, disorder-to-order transitions
  - Stable interactions form homo- and hetero-oligomeric protein complexes
- Conformational changes can alter PPIs (!)
- The interactome is the set of interaction in a give compartment

Driving force of protein-ligand interctions

- To maintain the order observed in biological systems, work is required
  - Protein folding requires work to be accomplished, since it requires a local entropy reduction
  - The synthesis of macromolecules requires the energy deriving from ATP hydrolsys
- The spontaneity of chemical reactions is expressed in terms of variation in Gibbs free energy (ΔG)
  - It is a state function: it does depend only on the initial and final state of the transition, not on

the path followed
  - It represent the energy of a system that can be used to produce work
  - It can be related to 3 other state functions: temperature (T), entropy (S) and entalpy (H)
    * $G = H - TS$
  - A decrease of G (a negative ΔG) indicates a spontaneous process
- Biological systems exist in a constant pressure and constant temperature environment: we can neglect changes in temperature
  - $\Delta G = \Delta H - T\Delta S$
- The entalpy of a system is a term that condenses its internal energy (E), volume (V) and pressure (P)
  - $H = E + PV$
- The internal energy E is composed of a kinetic and a potential term
  - $E = U + K$
- Biological systems are typically in a solid or liquid state: we can ignore volume and pressure changes
  - $\Delta H \approx \Delta E$
  - This is valid for bond formation/breaking, variation in weak interactions, variation in atomic motion indiced by heat
- Since we are at constant pressure, variations in entalpy equate heat transfer with the environment (Q)
  - $Q_p \approx \Delta E \approx \Delta H$
- When bonds are broken, energy is released: $\Delta H > 0$
- When bonds are formed, energy is absorbed from the environment: $\Delta H < 0$
- Entropy (S) is a quantity related to the number of microstates (Ω) that correspond to a given macrostate
  - $S = K_b \ln \Omega$
  - A microstate is a unique atomic configuration
  - A macrostate is an observable state of the system that can correspond to many unique atomic configurations
  - At constant temperature we observe that $\Delta S = \frac{\Delta H}{T} = \frac{Q_p}{T}$
- A process is at equilibrium when $\Delta G = 0$, endoergonic when $\Delta G > 0$ and exoergonic when $\Delta G < 0$
- Protein folding is spontaneous since the entropy reduction of the protein is more than offset from the entropy increase of the solvent
  - Entropy increase of water molecule of the solvent is a driving force in many biological processes
- Ligand binding is driven by the hydrophobic effect and Van der Waals interactions
  - PPI interfaces tend to be hydrophobic
  - The binding of small molecules is typically guided by geometry of the binding pocket and elecrtrostatic interaction
- The size of a PPI surface is related to the binding strenght
  - Standard interfaces are 1200 to 2000 Å^2
  - Low stability complexes have interfaces of 1150-1200 Å^2
  - Small molecules interact with proteins in a 300-100 Å^2 area, typically in deep pockets
- PPI interfaces tend to be flat and without pockets
  - The center of the interface tends to be particularly conserved
- Transient interfaces tend to be smaller, more hydrophobic and with better complementarity
- Homomeric interfaces resemble protein cores while heteromeric interfaces look more like non-binding surfaces
- The specificity of binding is given by electrostatic interactions
  - These also prevent aggregation
  - Specificity is low if a protein can bind many partners
  - PPI affinity ranges from pM to mM, but they are typically quite specific
- Some binding energies to be rembered
  - General range: -2.5 to -22 kcal/mol
  - Interactions in signal transduction are weak
  - Cofactor binding: -5.5 gto -9.5 kcal/mol
  - Antigen-antibody: -5 to -11 kcal/mol
  - Enzyme-inhibitor: -9 to -15 kcal/mol
  - Enzyme-transition state: -17 to -27 kcal/mol

- Interaction hotspots are PPI interfaces that are quite heterogeneous
  - The binding strenght is typically given by a few hydrophobic residues (hotspots)
  - They can be identified by alanine replacement: if we replace an hospot with alanine the affinity changes of at least 2 kcal/mol
  - Alanine replacement can also be done in silico (alanine scanning)
- Hotspots tend to account for less than 50% of the interaction surface, are quite conserved, appear in clusters
  - They are frequently represented by aromatic residues (F, Y, W)
  - They are surrounded by less important residues that shields them from the solvent

Protein docking

- It is biologically important to understand the molecular position of a ligand on a protein (its pose)
- Docking finds the optimal interaction, but cannot determine if the interaction actually happens or if it has a biological meaning
- Molecular docking is an optimization problem: we want to maximize the interaction of the ligand with the protein by operating on their torsion angles
  - We want to maximise electrostatic and geometric affinity
- We first perform a step called pose generation that explores the conformational space, and subsequently we rank the possible solutions
- A ligand can be a small molecule or also a macromolecule
- Docking is a golden standard in PPIs, because it is reliable, but it is computationally expensive: it is not an high throughput method
- A must for docking is to have a good 3d structure or model
- In order to test a docking method, I can take a structure with a co-crystalised ligand
  - I artificially separate the molecules and compare the docking prediction with the experimental data
  - This method is called bound docking
- On the contrary, in unbound docking it may be that I am using a structure for a protein that is in the wrong conformation for the interaction
  - A structure is defined native if in an uncomplexed state, pseudonative if complexed with a ligand different from the one used in the docking
  - I can also do unbound docking on a model
- Typical limitations of the apprach are: conformational changes, errors in the structures or models, limited computing resources