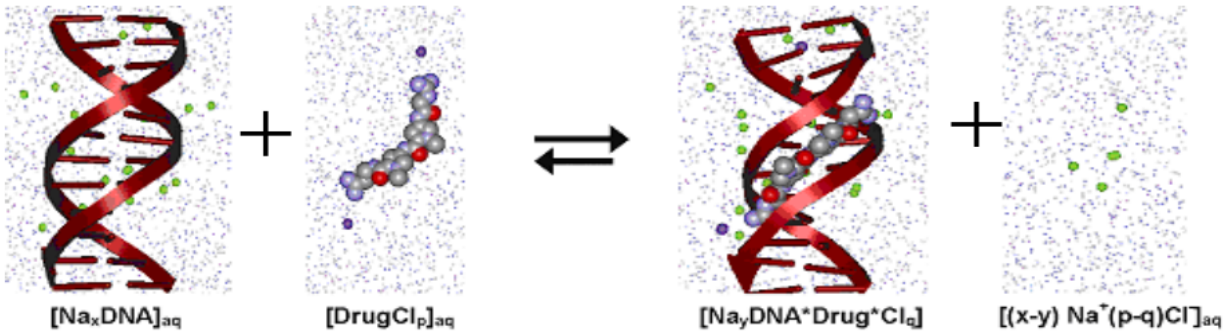
Forces involved in DNA-drug recognition:

Understanding the forces involved in the binding of proteins or small molecules to DNA is of prime importance due to two major reasons. Firstly, the design of sequence specific drugs having requisite affinity for DNA requires a knowledge how the structure of the drug is related to the specificity/affinity of binding and what structural modifications could result in a drug with desired qualities. Secondly, identifying the forces/energetics involved in such processes is fundamental to unraveling the mystery of molecular recognition in general and DNA binding in particular.

Some of the forces that are known to contribute to biomolecular recognition and also to DNA-drug binding are direct electrostatic interactions, direct van der Waals/packing interactions, complex hydration/dehydration contributions composed of hydrophobic component, solvation electrostatics, solvation van der Waals, ion effects and entropy terms.

DNA-drug binding may be described in the following manner,



Consider DNA-drug binding in an aqueous environment. DNA is polyanionic in nature and the drug molecule is also often charged. The associated counterions lie near the charged groups and are also partially solvated. When binding occurs, it results in a displacement of solvent from the binding site on both the DNA and drug. Also, since there would be partial compensation of charges as the DNA and drug are oppositely charged, some counterions would be released into the bulk solvent and are solvated fully. Also, the binding process would be associated with some structural deformation/adaptation of the DNA as well as the drug molecule in order to accommodate each other. All these events are associated with some energetic gains/losses, the comprehensive estimation of which is a major challenge.

Structural and conformational changes in the DNA and drug on binding in solution are associated with enthalpic and entropic contributions to the binding free energy, which can be theoretically estimated from ensembles of structures generated via simulations. The only drawback of this approach is the long time taken for the simulations.

The other terms, namely, electrostatics, van der Waals, hydrophobic component, rotational and translational entropy can be estimated from single structures.

PROTEIN – LIGAND INTERACTIONS:

Protein-ligand association

Proteinhavethefundamentalabilitytoselectively bindtoothermolecules.

Important for:

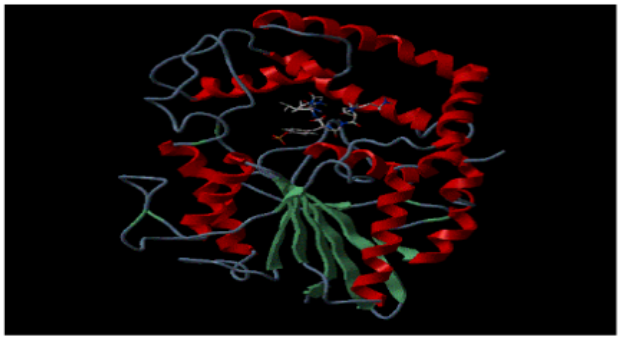
1.Enzymefunction.

2.Receptor actions (membrane).

3.Self-organization cellular structures and multicomponent protein complexes.

Importanttounderstand,bothquantitativelyand qualitatively.

Protein-ligand binding



$$P + L \Leftrightarrow PL$$

- Protein ligand binding is a spontaneous process.
- Similar forces as in protein folding are at work.
- Function of proteins is defined through its interactions with other molecules.

Overview

- Dissociation constant
- One or more possibly independent binding sites
- Cooperative effect:
  - The binding of the first ligand may affect the binding of the next ligand
  - Positive, negative cooperative effect
- Multivalent interactions:
  - Multiple possibly weak interactions between ligand and protein can lead to a strong affinity

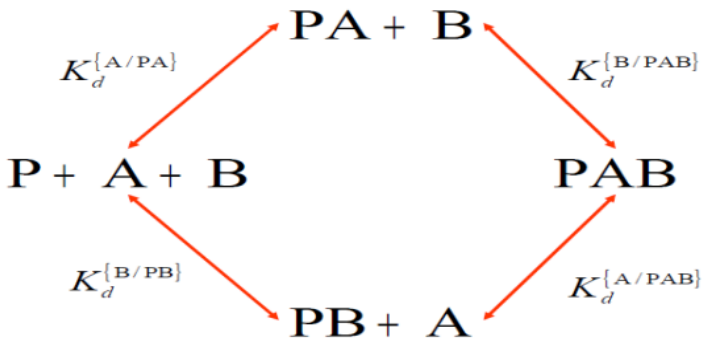
Single binding site

- Protein - ligand solution:
  - P, L, PL are given as concentrations
- Average number of ligand molecules bound to each protein:

$$\text{General: } \bar{n} = \frac{\text{concentration of L bound to P}}{\text{Total concentration of P}}$$

$$\text{Single binding site, one ligand species: } \bar{n} = \frac{[PL]}{[P] + [PL]}$$

Binding of different ligands to protein



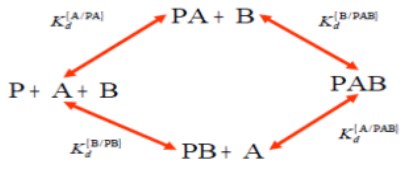
$$K_d^{\{A/PA\}} = \frac{[P][A]}{[PA]} \qquad K_d^{\{A/PAB\}} = \frac{[PB][A]}{[PAB]}$$

Independent binding

- Binding of A does NOT affect the binding of B and vice versa:

$$K_d^{\{A/PA\}} = K_d^{\{A/PAB\}}$$

$$K_d^{\{B/PB\}} = K_d^{\{B/PAB\}}$$



- Fractional saturation independent: $\theta_{AB} = \theta_A \times \theta_B$
- Independent binding can be treated as before.

Dependent binding

- Binding of A depends on the binding of B and vice versa:
  - Binding constants are different
- Positive cooperativity:
  - Binding of A (B) enhances (makes stronger) binding of B (A) $\Rightarrow$

$$K_d^{\{A/PA\}} > K_d^{\{A/PAB\}}$$
$$K_d^{\{B/PB\}} > K_d^{\{B/PAB\}}$$

Dependent binding

- Negative cooperativity:
  - Binding of A (B) makes the binding of B (A) weaker.

$$K_d^{\{A/PA\}} < K_d^{\{A/PAB\}}$$
$$K_d^{\{B/PB\}} < K_d^{\{B/PAB\}}$$

- Affects fractional saturation:
  - Positive cooperativity: Increase of [B] increase θ_A
  - Negative cooperativity: Increase of [B] decreases θ_A

Reasons for cooperativity

- Conformational changes induced by ligand binding
- (Un)favorable interactions between ligands.

PROTEIN – PROTEIN INTERACTIONS:

Pairwise protein–protein interactions:

cooperativity and organization:

Cooperativity is non-independence. Proteins are widely believed to fold cooperatively. Non-cooperative folding events would lead to an exhaustive search of the conformationa space to reach the global minimum. Yet, an exhaustive conformational search would imply time scales not affordable in the biological world. Considerable literature has addressed the challenging question of the physical basis of cooperativity through which proteins would avoid an exhaustive search. From the kinetic standpoint, cooperativity leads to preferred protein folding pathways. Cooperativity largely derives from the hydrophobic effect, the driving force of protein folding. Accounting for cooperativity has led to landmark experimental and computational investigations of the mechanisms and pathways of protein folding, addressing the question of *how* the protein chain searches the immense number of possible nonlocal interactions to yield the hydrophobic core. To understand cooperativity, we need to think of the system as a cohesive unit, where the behavior of the parts may depend on each other. That is, the overall behavior is the outcome of the properties of the entire system and not of the sum of the properties of its components. Below, we argue that the thermodynamic stability of the protein–protein complex is not a summation of the individual contributions of each of the residues independent of the other; rather, residues which are in direct spatial contact, or in close contact through a few tightly packed intermediate residues, impact the stability of the association in a non-additive manner. When a residue is in a tight physical (chemical) geometrical contact with others, its substitution would affect the structure and interactions of its neighboring residues. Thus, if this residue and its neighbors contribute significantly to the stability of the molecule or the complex, its mutation may affect the stability not only through the change of its own interactions, but in addition through the changes of its neighbors. This would affect the stability of the complex beyond the direct altered interactions of the mutated residue. Hence, if we were to simultaneously mutate two residues which are in close spatial contact in a densely packed environment, the change in the stability would not be the sum of the measured changes of each one separately. The measured change in the thermodynamic stability upon a mutation of a single residue already takes into

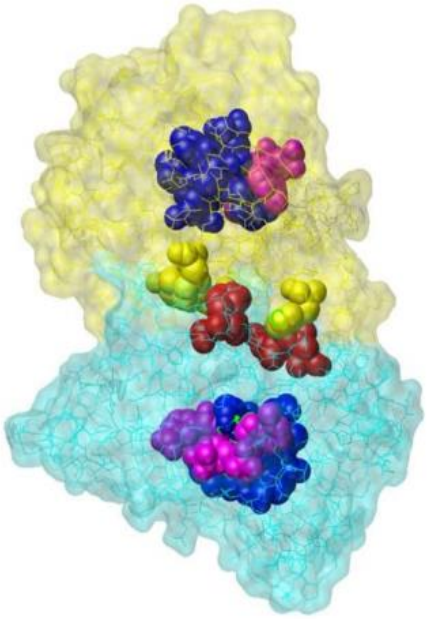
account changes in the interactions of its closely packed neighboring residues. On the other hand, if the protein–protein interface can be separated into cohesive separate units, the impact of mutations in each of these is independent, i.e., non-cooperative.

Protein–protein binding sites consist of independent regions

Here, we propose that in protein–protein complexes, the binding sites consist of one or a few independent, tightly packed regions . Residues which contribute significantly (more than 2 kcal mol−1 to the free energy of the protein–protein association are clustered within these regions. Within the tightly packed cluster, these so-called hot spot residues form a network of interactions, thus contributing *cooperatively* to the stability of the protein–protein complex.

In contrast, the contribution of the independent regions is additive. We name these regions 'hot regions'. This type of tightly packed organization effectively screens the solvent from the charged groups, strengthening the charge–charge interactions. This rationalizes why the hot spots do not form more salt bridges and hydrogen bonds than other interacting residues at the protein–protein interface. Such a *hot region* organization is advantageous to the protein associations. Moreover, this organization of protein–protein binding highlights the similarity between protein folding and protein binding In both packing plays a crucial role. Within the packed protein cores and the packed hot regions, there are residues that contribute significantly and cooperatively to the stability. As in cores, the hot spots arealso highly conserved by evolution at the protein binding site. The average conservation ratio of the neighboring residues in hot regions is 0.47 as compared to a 0.26 average conservation ratio for the rest of the interface residues. A residue is identified as a 'hot spot neighbor' if the distance between its Cα and a Cα of a residue is less than 6.5 A° .We also observe that the packing is higher around hot spots (on average 7.0 residues in the hot regions) and lower at the other regions (5.6 residues outside). For the packing calculations, residues whose Cα are closer than the cut-off distance are defined to be in contact, excluding the two bonded sequential neighbors. These numbers are obtained from a set of 44 interface clusters . Combined, the hot region organization provides a new description of protein binding sites. It is useful since it explains why a summation of the hot spots contributions over-estimates (or, under-estimates) the binding free energy. This is expected to lead to better scoring schemes in the prediction of protein–protein associations. In addition, it suggests that a hot region should provide a good target for drug design. Figure presents an example displaying the

analogy of hot regions in protein interfaces with protein cores. The light yellow and cyan are the two chains of a protein, magenta atoms belong to the NADPH molecules, dark blue atoms represent the folding core of the protein. Dark yellow and red atoms indicate the hot region in the interface between the two chains. Both the folding core and the hot region have similar organization of highly packed clustered atoms.



The new definition of a protein–protein binding interface: an interface is comprised of cooperative, locally densely packed 'hot regions'. This view of a cooperative hot region is attractive since it is consistent with current understanding of cores in protein folding. In agreement with the notion that protein folding and protein binding are similar processes with similar underlying mechanisms, a cooperative 'hot region' may resemble a core of a domain. The absence of cooperativity between hot regions may conceptually be viewed as two stable domains in a multi-domain protein. Hence, binding and folding are similar: cooperativity is observed in local tightly interacting regions and in the protein cores. Sheer counting under-represents the number of conserved charged residue couples in binding and in folding. In both interacting across-the-interface hot regions and in protein cores, electrostatics is enhanced through solvent screening.