Protein-Ligand Interactions: “Locks-and-keys”?

R. Sowdhamini
NCBS (TIFR)

14 December 2007
Protein-Ligand Interactions

• The importance of protein ligand binding in biological systems should **not be underestimated**. Any organism must have a mechanism of interacting with its environment. Individual cells must be able to interact with a complex variety of different molecules, derived from not only the outside environment but also generated within the cell itself.

• Protein ligand binding has an important role in the **function of living organisms** and is one method that the cell uses to interact with these wide variety of molecules. In order to understand this in more detail, it is best to first define what we mean by a **ligand**.
What is a ligand?

- A ligand is simply a molecule that interacts with a protein, by specifically binding to the protein.
- Technical definition: A molecule that binds noncovalently and specifically to another, A substance that binds to a protein or other molecule noncovalently and specifically.
- A ligand can be a nucleic acid, polysaccharide, lipid or even another protein.

Protein ligand binding is involved in many cell functions including hormone receptors, gene regulation, transport across membranes, the immune response and enzyme catalysis.
Protein-Ligand Interactions

- Binding is specific to a particular ligand or group of ligands (allostery)
- Binding occurs at a particular site in the protein molecule
- Binding is reversible
Types of protein-ligand interactions

1. Reversible Binding: Single ligand
2. Reversible Binding: Two ligands, Allosteric binding
3. Reversible Binding: Two ligands, Competitive binding
Lock-and-key hypothesis

- The lock and key hypothesis is an early theory introduced by Emil Fischer in 1894. It basically states that the enzyme is the lock and the substrate is the key. Only a specific shape substrate would fit into its corresponding enzyme.

- A hypothesis highlighting that enzymes and substrates have very specific intermolecular interactions was proposed much before structures were known for these molecules and their interactions.
112 years ago, Emil Fischer proposed this descriptive and provocative analogy for molecular recognition: the lock and key hypothesis. At a time when little was known of the molecular structures of even the relatively simple substrates of enzymes, let alone the complex structures of proteins, this gave an extraordinarily useful visual image of enzyme action.
Similar recognition processes, such as antigen-antibody, hormone/growth factor-receptor, lectin-sugar, repressor-DNA, have since been identified in other classes of proteins. Can the Emil-Fischer be applied to these systems? Has the hypothesis stood the test of time?
Availability of crystal structures of proteins and complexes ...

• A few of these complexes are of relatively rigid ligands; but even these generally show some conformational accommodation of the protein during recognition/binding.

• In many complexes, there is evidence of major conformational changes, or even hinge bending in multi-domain proteins, so that the ligand is tightly bound to the substrate and often removed entirely from the aqueous environment.
Availability of crystal structures of proteins and complexes ...

• Even if the proteins and ligands are flexible, the complex most involves a well-defined complementarity of the bound components; this is not dissimilar to the original ‘lock-and-key’ hypothesis.
Recognition modules: lectins

- Binding modules or domains that are used for recognition in multimeric assemblies that mediate cell-cell interactions and act as adaptors to bring together other molecules.
Recognition modules: lectins

- Each member in the superfamily comprises an elaborated jelly-roll fold. Identical hydrogen bonding patterns interconnect a certain number of anti-parallel strands, although they are much longer for lectins.
- The resultant structure is a sandwich, in which one sheet is relatively flat and the other strongly concave – where the sugar binding happens. This provides a rigid framework for binding the ligand.
- Carbohydrate and calcium binding
Recognition modules: lectins

- Adaptor domains mediate interactions by forming a rigid framework after oligomerisation.
- Other adaptor domains, like SH2 and SH3 also achieve similar rigid framework.
- Substantial specificity is achieved at the concave surface.
Proteinases: flexible ligand

• A polypeptide is a flexible ligand with a very large number of low energy conformers energetically possible.
• Of these, only those that allow access to both sides of the scissile peptide are useful.
Proteinases

• Most proteinases – like serine, aspartic and matrix metalloproteinases – involve a distorted strand structure involving parallel or antiparallel sheet interactions.
Proteinases: hydrogen bonding between substrate analogue and enzyme

- Hydrogen bonding with main-chain function occurs on both sides of the scissile bond.
- These interactions allow a generic mode of binding for a variety of sequences, but a very precise orientation of the scissile bond with respect to catalytic aspartates.
Proteinases: flexible ligands

- On the N-terminal side, the main chain H-bonding is replaced by a threonine gamma-hydroxyl to NH of P3.
- Transition state is stabilised by “charged H-bonds”.
- The specificity is achieved by complementarity between side chains in the enzyme active site: these involve van der Waals interactions (types of atomic interactions).
Proteinases: differences in specificity

- Most aspartic proteinases retain a “flap” loop that closes-in a conserved Tyr75 contributes to the specificity pocket.
- In addition, renin, cathepsin D and yeast proteinases retain an additional flap (193-203), involving polyproline loop that interacts with the ligand at P3’ and P5’ positions.
Proteinases: flexible ligands, better inhibitors

• Identical side chains of inhibitors can have radically different positions and conformations and can opportunistically bind better.

• This emphasises the need to study a large number of interactions if predictive success is to be achieved in drug discovery.
Recognition and bending through hinge-bending

• When a ligand needs to be entirely surrounded by the protein, hinge-bending between domains is often done to achieve best fit: examples are NAD-binding proteins like dehydrogenases, maltose binding proteins and anion binding proteins like phorphobilinogen deaminase.
Anion-binding proteins

- Doubly-wound fold with four parallel and one anti-parallel beta-strands.
- In anion binding proteins, the ligand is at the C-cap of beta-strands and N-cap of helices.
- Extensive hydrogen bonding delocalises the charge of the ligand and stabilises the interactions.
Phorphobilinogen deaminase

- This enzyme has a pyrrole as a primer that is covalently attached to the enzyme.
- The enzyme assembles the primer and then adds four pyrrole rings by successively deamidating phorphobilinogen.
- The tetrapyrrolole is cleaved off to form a haem, chlorophyll or vitamin $\text{B}_{12}$. 
Phorphobilinogen deaminase: Protein-ligand interactions

The tetrapyrrole is stabilised primarily by ionic interactions with the protein.

Several positively charged residues interact and the side chains of the substituted tetrapyrrole.
Flexibility in phorphobilinogen deaminase

• In another superfamily member, transferrin, hinge bending has been noticed between third and fourth beta-strand.

Hinge bending between domains allows access to the substrate at each stage in the polymerisation. It also allows the developing polymer to be accommodated in such a way that the same catalytic apparatus can be used at each step in the reaction.
Flexibility in proteins: Staphylococcal Adhesin Binding to Fibrinogen

SdrG is a cell wall-anchored adhesin from *Staphylococcus epidermidis* that binds to the B chain of human fibrinogen SdrG of (Fg) and is necessary and sufficient for bacterial attachment to Fg-coated biomaterials.
Flexibility in proteins: Staphylococcal Adhesin Binding to Fibrinogen

Structure comparison of apo-protein and ligand-bound form.
Flexibility in proteins: SdrG binds to its ligand with a dynamic “dock, lock and latch” mechanism.

- Apoprotein contains a putative ligand binding cleft between the N2 and N3 domains and an empty latchthing cleft between D and E strands that can accommodate a guest strand.

- This binding mechanism involves the docking of the ligand in a pocket formed between two SdrG subdomains followed by the movement of a C-terminal extension of one subdomain to cover the ligand and to insert and complement a β-sheet in a neighboring subdomain.
Flexibility in proteins: SdrG binds to its ligand with a dynamic “dock, lock and latch” mechanism.

• Biochemical studies (FRET) and engineering of a disulphide bond, there is now a proof that ligand binding precedes structure compaction: the stable-compact structure does not accept ligand and ligand binding is restored with addition of DTT.
Molecular recognition by multimeric assemblies

• In the 1970s, inspired by the structures of haemoglobin etc., there has been a theory that signalling is edited by ligand binding and that long-distance transmissions are possible by steric changes across subunits and largely due to allostery.

• However, since the 1980s, it has been realised that proteins, in fact, form assemblies and aggregation rather than allostery. Standing examples are growth factor-receptor interactions.
Growth factor-receptor interactions: receptor closes-in on ligands

• The structures of complexes of growth hormone and TNF show that the extracellular domains of the membrane bound receptors assemble around the ligands, rather than the ligands docking into pockets or clefts of the receptors.

• This presumably brings the helices of the ligands to higher proximity.
Biological introduction: Domain architecture of KIT

• KIT is composed of a glycosylated extracellular ligand-binding domain (ectodomain) that is connected to a cytoplasmic region by means of a single transmembrane (TM) domain (reviewed in Schlessinger, 2000).

• The ectodomain of KIT and other members of type III RTKs all contain five Ig-like domains, (designated as D1, D2, D3, D4, and D5) in which the second and third membrane distal domains were shown to play a role in ligand recognition (reviewed in Ullrich and Schlessinger, 1990).
Biological introduction: Domain architecture of KIT

- The cytoplasmic region of KIT contains a protein tyrosine kinase (PTK) domain with a large kinase-insert region, another hallmark of type III RTKs.
Crystal structure of ectodomain of KIT
Crystal structure of ectodomain of KIT in complex with SCF dimer

The structure of SCF-KIT complex shows 2:2 stoichiometry, in which two sets of 1:1 complexes in the asymmetric unit are related by a noncrystallographic 2-fold symmetry.
Sites of interaction
Sites of interaction
Structural Changes upon ligand binding: D1, D2 and D3 domains

• Superimposition of the structures of individual D1, D2, and D3 of KIT monomeric form with corresponding structures of the SCF-induced homodimeric form reveals rmsd values of 0.5, 0.8, and 1.1 Å° for 82, 92, and 100 aligned Ca residues in D1, D2, and D3, respectively. Similarly, superimposition of the structure of the entire D1-D2-D3 region of KIT monomers with the corresponding structures in the SCF-KIT 2:2 complex reveals an rmsd of 1.1 Å° for 274 aligned Ca residues of the D1-D2-D3 region.
Large rearrangements happen at D4 and D5 domains upon ligand binding

By contrast, superimposition of the D3-D4-D5 region of KIT monomeric form with the corresponding region in the homodimeric form reveals a large structural change in the orientation of D4 and D5 relative to each other and relative to the ligand binding region of KIT.
The homodimeric interface at D4 and D5 are kept largely electrostatic (in fact of like charges) so that not only the receptor will be stable as a monomer but will repel each other until the ligand binding induces the receptor to overcome these forces and miraculously the homodimer at D4 and D5 would be stabilised by salt bridges after a small rotation (15 degrees) of these two domains (Arg381 and Glu386’).
Inositol triphosphate receptor domain arrangements are more compact via flexible linkers in the presence of Ip3.
Flexibility in proteins

• Most proteins use flexibility of their internal structure to provide a complementary surface by induced fit. Alpha-proteins or alpha-beta-proteins provide more malleable structures where helices can move with respect to other helices or parallel beta-sheets.

• Further, to this loops between secondary structures can move to permit access to ligands and subsequently bind over the ligand.
Proteins can trap ligands away from an aqueous environment: ammonia transporters

- The ammonium transport proteins (methylamine permeases ammonium Transporters rhesus) are present in all domains of life; however, functional studies with members of this family have yielded controversial results with respect to the chemical identity (NH4 or NH3) of the transported species.
The structure of wild-type AmtB from *Escherichia coli* in two crystal forms at 1.8- and 2.1-Å resolution, respectively. Substrate transport occurs through a narrow mainly hydrophobic pore located at the center of each monomer of the trimeric AmtB. At the periplasmic entry, a binding site for NH₄ is observed. Two phenylalanine side chains (F107 and F215) block access into the pore from the periplasmic side. Further into the pore, the side chains of two highly conserved histidine residues (H168 and H318) bridged by a H-bond lie adjacent, with their edges pointing into the cavity. These histidine residues may facilitate the deprotonation of an ammonium ion entering the pore.
• Residues from helices TM1, TM3, and TM5 and their pseudosymmetry mates TM6, TM8, and TM10 form the entries and walls of the blocked pore.

• The two histidines, H168 and H318, are located at the N termini of helices TM5 and TM10, respectively, and are related by the pseudotwofold symmetry. They are preceded by two highly conserved aspartates, D160 and D310.
One more example of free ammonia: hydrophobic channel

- Glucosamine-6-phosphate synthase catalyses the first and rate-limiting step in hexosamine metabolism, converting fructose 6-phosphate into glucosamine 6-phosphate in the presence of glutamine. The crystal structure of the Escherichia coli enzyme reveals the domain organisation of the homodimeric molecule. The 18 Å hydrophobic channel sequestered from the solvent connects the glutaminase and isomerase active sites, and provides a means of ammonia transfer from glutamine to sugar phosphate.

- Based on the structure, a mechanism of enzyme action and self-regulation is proposed. It involves large domain movements triggered by substrate binding that lead to the formation of the channel.
• Ammonia produced by glutamine hydrolysis should remain trapped in the enzyme to avoid protonation and to retain its nucleophilic character.

• The crystal structure of GlmS presents an internal channel between the two active sites, with a separation of 18 Å.
• The structural data and energetic considerations strongly indicate that the methylamine permeasesammonium transporters rhesus proteins are ammonia gas channels.
Lock-and-key hypothesis: does it hold good?

- There are few systems that even approximate to this hypothesis. The adaptor proteins (lectins, SH2 and SH3 domains), with their tight hydrogen-bonding network within their beta-sheets, seem to come close to rigid systems. Even, they do seem to have loop differences to achieve different oligomerisation modes and ligand binding.
Lock-and-key hypothesis: does it hold good?

- However, the precise mode of interactions between proteins and ligands are lot like the lock-and-key hypothesis.
- Since conformational changes are hard to predict, given the structure of the uncomplexed form of the protein, the rigid lock-and-key hypothesis will remain the working model for many pharmaceutical, biotechnology industries for the design of ligands.