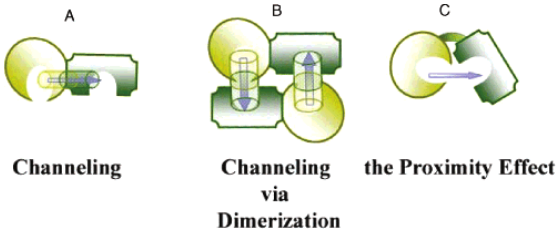
Biochemistry Lab

Introduction to Protein-Protein Interaction

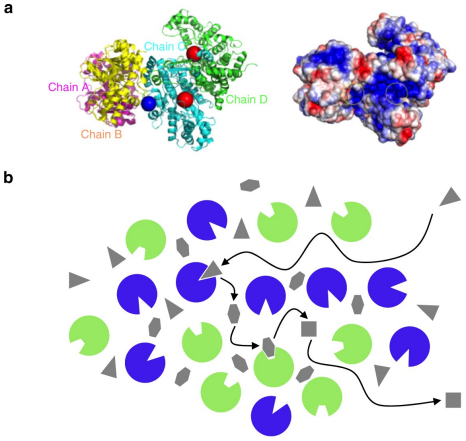
Introduction



Taken from Singh & Christendat *Biochemistry* 45, 7787 (2006)

In any given cell, there are tens of thousands of proteins performing a large range of biochemical and physiological functions. Some of these processes are involved in sequential metabolic pathways, proteins which bind and control the functions of other proteins, binding in complex formation with carbohydrates, lipids and nucleic acids and proteins forming structural elements within and around cells. Some proteins function independent of others while scientists have long known about strong assemblies of proteins performing important functions. Regardless of the type of protein involved in these interactions, a common feature is that proteins will dynamically interact or bind to many compounds including other proteins. Impact of protein-protein interactions include the regulation of kinetic properties/allosteric regulation of enzymes, structural shifts of a protein induced by the binding of a second protein which could alter the binding site for effector molecules or active sites. Protein-protein interactions also shift the cellular location of one of the binding partners controlling its activity.

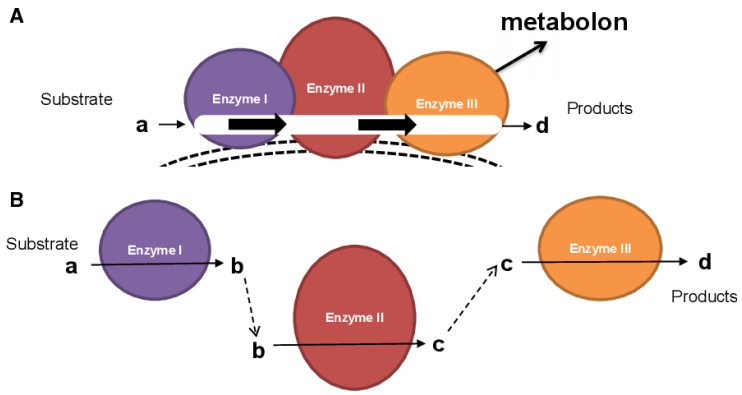
A metabolon is a specific grouping of proteins binding to other proteins forming a complex of enzymes allowing sharing of reactants and products. This was initially described as "substrate channeling" by Dr. Paul Srere in 1985. In these early studies of sequential interaction of proteins of carbohydrate metabolism many interactions were identified. These metabolons or supramolecular complexes of sequential metabolic enzymes were thought to interact due to a lack of available bulk water. It was understood that many cells and especially certain organelles like the mitochondria did not have enough water to solvate the polar surfaces of the proteins, lipid headgroups and other metabolites. A simple aggregation of proteins would "solvate" each other, decreasing the surface area needed for water-protein solvent systems. For these metabolons, a second benefit was the potential substrate channeling. Such channeling could occur directly as proposed for MDH and CS or by location of a dimer of dimers with active sites facing each other or due to the close proximity of each other increasing the local concentration of metabolites.



Mechanisms of substrate channelling in dynamic enzyme assemblies. **a** Direct channelling by electrostatic retention of the channeled metabolite on the surface of the enzyme complex. A structural model of the bovine malate dehydrogenase (MDH)–citrate synthase (CS) complex is shown. On the left, the polypeptides are illustrated as ribbon diagrams, with the MDH dimer shown in magenta and yellow and the CS dimer show in green and cyan. The blue circle shows where OAA molecules were initially placed in a Brownian dynamics simulation. Red circles show the active sites in the CS dimer. On the right, the surface structure of the complex is shown, with red and blue colors representing negative and positive electrostatic potential, respectively. Neutral regions are shown in white. Yellow circles indicate the positions of the adjacent MDH and CS active sites. **b** Probabilistic channelling within a large cluster of enzymes. Two enzymes are shown as green and blue circles. Metabolites are shown as grey polygons, with each shape representing a different metabolite. The arrows indicate the path taken by metabolites in a sequential conversion event by two enzymes. Taken from Sweetlove & Fernie. Nature Communications 9, 2138 (2018)

Organized systems of sequential proteins in a common metabolic pathway would result in a local concentration of shared products of one enzyme which in turn would be the reactant for another enzyme. This local enrichment would support increased reaction rates without a need for random diffusion of metabolites, isolation of metabolic intermediates from competing side reactions, limiting the half-life on unstable product/reactant pairs and the sequestering of possible cytotoxic metabolites. Some protein complexes are well known including the

pyruvate dehydrogenase complex. The initial work of Srere and others focused on the tricarboxylic acid (TCA, AKA Krebs cycle) in mammalian systems. Since then other organisms have been identified to have similar TCA metabolons including plants. While the existence of these metabolons are now well founded, the search for specific interactions, the critical binding determinants and their regulation extents to many systems across all organisms. Non metabolic protein complexes are termed protein "interactomes" and are in control of many important physiological processes in normal cell function and when dysregulated often are driving forces of diseases. As a result many new previously thought "undruggable" targets are thought to have a great potential as an intervention target for novel drugs modulating protein-protein interactions.

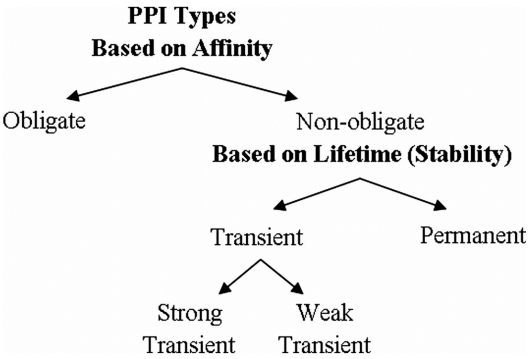


The substrate channeling association. Metabolic pathway in which product d is synthesized from substrate a via the reactions catalyzed by enzymes I, II, and III (arrows). b and c are pathway intermediates. **(A)** Association of the metabolon will enhance pathway reactions and is expected to upregulate the pathway. **(B)** Dissociation of the metabolon will downregulate the pathway. Zhang and Fernie. Plant Commun, 2 100081 eCollection (2021)

Types of Protein-Protein Interactions

The physical interaction between proteins "docked" to another can be weak and unstable or transient in nature or composed of many weaker non-covalent interactions providing a durable pairing of proteins. The interface between docked proteins requires both shape and chemical complementation. These are the forces that help define types of protein-protein interactions.

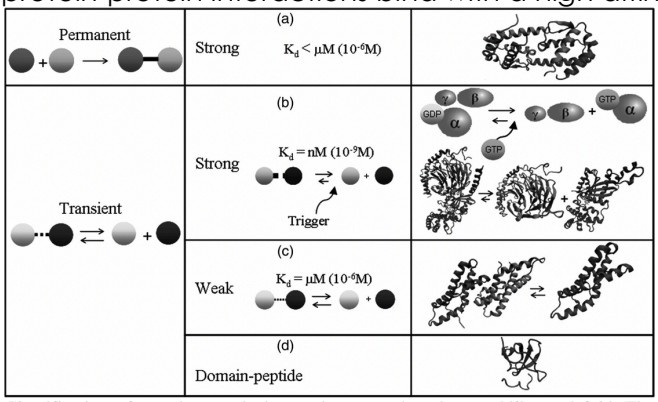
Strong or weak, protein-protein interactions have been organized based on their composition (one protein or more), affinity (binding association) and the life-time or duration of the interaction. Homo or hetero-oligomeric complexes interactions are pretty obvious. Both MDH and CS are two separate protein chains held together as a homodimer. These are examples of homo-oligomeric complexes. The combination of MDH and CS is of course a hetero-oligomeric complex. It is interesting that many multi-protein complexes use a stable homo-dimer to form a larger hetero-protein complex. Proteins that do not function or are unstable as monomer, that is without their protein interaction partners, are considered obligate complexes. Obligate in science is used as a term to describe "by necessity". Thus proteins involved in a protein-protein interactions that are unstable or will not function without the binding partner are required obligate interactions. These types of proteins are often referred to as quaternary structure of an oligomeric protein. Non-obligate interactions are simply the opposite. For example, MDH and CS can function as monomers or as interacting complexes, and thus are non-obligates. These proteins that can survive alone or as a multimer are further characterized based on the half-life of their interaction.



Relation of protein–protein interaction types based on affinity and stability. Non-obligate interactions are transient but there are some examples of permanent non-obligate interactions such as enzyme–inhibitor interactions
Protein Eng Des Sel, 24, pp 635–648 (2011)

Some proteins are considered permanent due to the irreversible nature of their interaction. Transient complexes have a wider range of binding association/disassociation (Ka/Kd; affinity) classified as strong or weak transient interactions. Strong transient interactions are often involved in signaling processes with

a classic example being GTP binding proteins involved in receptor signaling. While permanent obligate protein-protein interactions bind with a high affinity often in the micromolar range or lower, transient interactions range in the mid-high micromolar range or greater. Interesting some of the more strong transient interactions can interact with dissociation constants (Kd) in the nanomolar range but only after some triggering event takes place. Examples include phosphorylation, small regulating compounds or other post translational modifications including methylation and acetylation. Proteins with intrinsic disordered domains also can form transient protein-protein interactions as these domains create and unmask potential protein docking sites. There are notable structural differences between transient and permanent protein-protein interaction sites. Transient interactions often involve helix and turns which provide flexibility required for binding and release from the complex. Whereas the obligate permanent interaction interfaces often include helix and beta sheets, which combine to create a more stable and fixed surface for protein binding. The interfaces of transient interactors will often be polar/charged forming salt bridges along with residues with potential hydrogen bonds are likely to be involved in these types of interactions. Obligate permanent protein-protein interactions have a higher potential for hydrophobic interactions. These are trends and no one type of non-covalent interaction is found only with one type of protein complex.



Classification of protein–protein interaction types based on stability and fold. The mechanisms of association and dissociation processes are shown for strong and weak transient protein–protein complexes, along with the structures of example cases. (**a**) Permanent protein–protein interaction: Components are stable only in complex form (**b**) Strong transient protein–protein interaction: association/dissociation takes place under certain triggers such as chemical modification, conformational change and colocalization; dissociation constant (Kd) is in nanomolar range, (**c**)Weak transient protein–protein interaction: Complexes are broken and formed continuously and Kd is in micromolar range, (**d**)Domain–peptide interaction: a globular domain recognizes a short linear motif ***Protein Eng Des Sel.* 24, p 635–648(2011)**

Kinetic and Reaction Coordinates of Protein-Protein Interactions

The act of binding of two proteins can be thought of as a kinetic event. Two free (non-bound) proteins A and B will bind forming a transient hetero-oligomeric protein-protein complex. The equilibria constants describe the association (K_a) and disassociation (K_d) of two proteins. The strength or binding affinity of a protein-protein interaction is the K_d which is defined as the ratio between rate constant of the complex dissociation reaction (off rate: K_{off} or K_d) and that of the association reaction or the on rate (K_{on} or K_a; scheme (1)). K_d is a concentration dimension and as expressed in the equilibria equation, if the K_d is high, the reaction tends to proceed in the reverse, that is the protein interaction easily dissociates, has a low affinity, and the ratio of bound (AB) to free (A and B) form is low. This is the description for simple two state (free and bound) interactions.

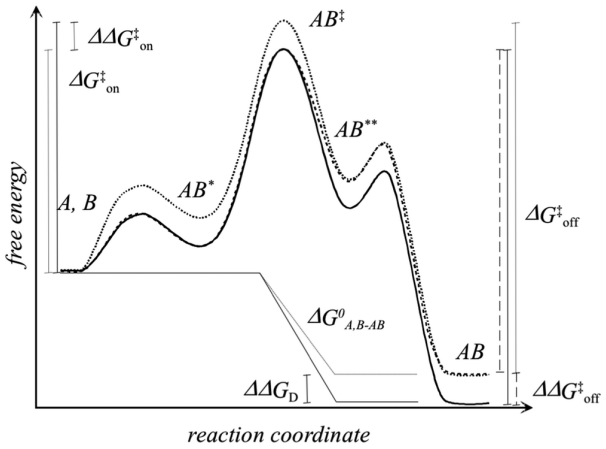
$$A + B \underset{k_d}{\overset{k_a}{\rightleftarrows}} AB; \quad K_d \equiv \frac{k_d}{k_a} \tag{1}$$

$$A + B \underset{k_{-D}}{\overset{k_D}{\rightleftarrows}} A * B \xrightarrow{k_c} AB; \quad k_a = \frac{k_D k_c}{k_{-D} + k_c} \tag{2}$$

$$A + B \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} AB^* \underset{k_{-2}}{\overset{k2}{\rightleftarrows}} AB^{**} \underset{k_{-3}}{\overset{k3}{\rightleftarrows}} AB \tag{3}$$

There are other binding mechanisms that include an temporary form (A*B). The three state kinetics assumes a intermediate or transient complex is part of the binding process. The transient complex can either dissociate back into the lone free proteins or form the final bound pair (AB). The overall rate constant of association k_a is dependent on diffusion-controlled rate constant (k_D), dissociation rate constant of the transient complex (k_{-D}) and the conformational rearrangement rate constant (k_c) as shown in scheme 2. A second description of proteins forming a complex involves an unstable binding complex (AB*) which can form into a more stable, and slightly different (AB**) complex. Each complex is describing the shape of each protein and the bonds formed between proteins during the interaction until the final form of the protein-protein interaction where all the final bonds are formed for the protein complex.

In terms of reaction coordinates, a protein complex is formed if the free energy of the bound state is lower than the free energy of the unbound proteins. For a second order reaction, the free energy profile is concentration-dependent. The association first meets the other while bound to surface waters (solvent encounters) and as interactions between residues of opposite proteins begin to take place the waters associated with H bonding/polar/charged residues are replaced with the non-covalent interactions between proteins. The process of a protein being fully solvated and non-interacting and then the final interface where many of the solvent water molecules have been replaced by protein-protein interaction can happen at one time or in several states (intermediates) as described by the three kinetic schemes and the reaction coordinates. Dissociation goes through the same transition state, and is characterized by breaking of the short-range interactions between two proteins.



Free energy profile describing the pathway for the formation of a protein–protein complex. Formation of the complex (AB) from the free proteins A and B *via* the encounter complex AB*, the transition state AB‡ and the intermediate AB**. Comparison of the profiles for the wt proteins (—) with a mutant affecting long-range electrostatic interactions (·····) and a mutant affecting short-range interactions (----), respectively. The free energies ΔG are indicated both for the complex formation ΔG° and the transition state, as well as the changes in free energy of the encounter complex. G. Schreiber, CHAPTER 1:Protein–Protein Interaction Interfaces and their Functional Implications, in *Protein–Protein Interaction Regulators*, p. 1-24 (2020)

Forces Influencing Protein-Protein Association Rate

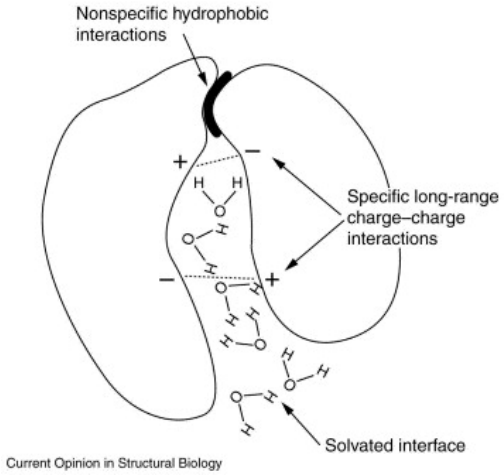
Factors that can alter the rate of association include solvents/co-solvents, pH and ionic strength of the medium. The association of proteins to engage in interactions requires diffusion of the two proteins to collide with the correct geometrical orientation. This is a slow step, especially in purified systems free of the native conditions found inside of a cell. Perturbing the diffusion step is possible by increasing the viscosity of the solution used to test interactions. In fat, the *in vivo* concentration of biomacromolecules in the cytoplasm is estimate to range between 50 and 400 mg/ml with ~70% of this due to proteins. Such additional solutions include using glycerol or sucrose to reduce diffusion. In some studies, the relative association rate was found to be linearly dependent on the relative viscosity of the medium. Crowding agents are solutions that tie up water mimicking the crowded environment inside a cell or organelle. Crowding agents include polyethylene glycol (PEG) dextran and Ficoll. Inert or non-interacting proteins such as bovine serum albumin is also often used as a crowding agent. Macromolecular crowding agents differ from viscosity agents as crowding agents are long polar/hydrogen bonding polymers that take space, effectively creating an excluded volume. That is the volume of a solution (inside the cell) that is inaccessible to other molecules due to the space taken up by the crowding agent. Thus these agents minimize free water, increase viscosity, and increase the effective concentration by making the effective volume smaller. In less simple terms, crowding agents stabilizes the native state of a protein-protein interaction by destabilizing the unfolded state, thus compensating for the energetically unfavorable folded and free protein conformation. Salts also impact protein-protein interactions as they can mask electrostatic attraction between interacting residues. These ionic interactions between proteins are a driving force for fast association. pH can alter the interaction rates of two proteins. Some proteins, whose protein interfaces involve histidine are often impacted by pH of a solution. Because histidine is the only amino acid with a pKa near the physiological pH, shifts in pKas of histidine along with the pH of the solvent play a critical role driving and sustaining interactions.

Forces Involved in the Interface of Protein-Protein Interactions

The interface between two proteins involved in protein-protein interaction requires both shape and chemical complementation. That is both the fit and the forces that keep the proteins together must have a tight match. The forces driving these interaction include residues which present opportunities for

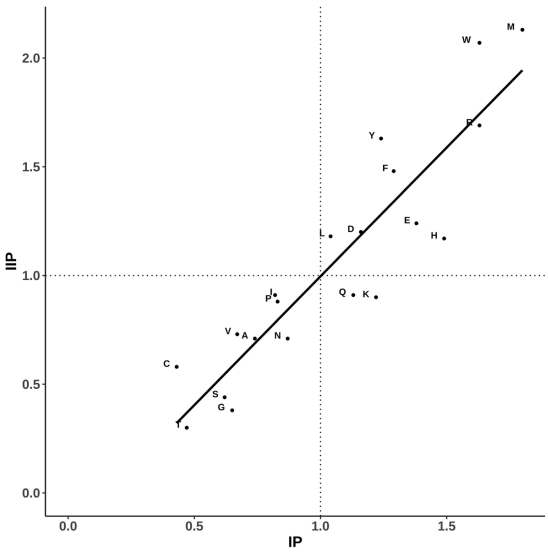
salt bridges, hydrogen bonding, hydrophobic effects and van der Waal contacts. While weak forces, they combine to form an interface for specific molecular interactions. In addition to the polar, hydrophobic, charged interacting residues which make up about 70% of the interface, most interaction interfaces include a significant occupation of water molecules.

Most of the binding energy proteins can be attributed to a number of "hot-spot" residues. These amino acids are defined as those that, upon mutation to alanine, lose 1-2 kcal/mol binding energy. Usually protein-protein interactions happen on several key amino acid residues in the interaction region. Tryptophan, arginine and tyrosine are more likely to appear in hot-spots than other amino acids. Because of the importance of these binding determinants, they have become a target for creating small molecule drugs or peptides to act in a specific manner on these hot-spots.



The transition state for association. Specific, long-range electrostatic interactions and nonspecific hydrophobic and van der Waals interactions stabilise the transition state, with large parts of the interface being solvated. Current Opinion in Structural Biology Volume 12, Issue 1, 1 February 2002, Pages 41-47

A detailed study of amino acid residues in the interface of interacting proteins using known 3-D structures of protein-protein complexes was performed to show the role of these residues in and out of their complexes. The scientists measured the strengths and types of interactions each amino acid has in and out of the interface of a protein-protein complex. Those amino acids that interacted within their own protein (intra bonding) and amino acids of the partner protein (inter bonding). Many of the residues that interact with amino acids of the partner protein also interact with amino acids or main-chain of the residue's own protein (inter and intra bonding) and only a quarter of the interactions identified only interacted in an inter bonding fashion. This can help give insight to one while searching structures of docked proteins – look for amino acids not just close to the opposite protein, but don't overlook the residues that might bind to their own protein and the partner protein. Residue types with good propensity for simultaneous intra and inter protein interactions include hydrophobic residues Leu, Phe, Trp and Met and also polar residues Asp, Glu, His and Arg. Therefore, the simultaneous intra and inter-protein interactions include various kinds, such as interactions between hydrophobic groups and hydrogen bonds.



Scatter plot showing the propensities of the residue types to occur in the protein-protein interfaces (IP, along the X-axis) and propensities to form simultaneous inter and intra-protein interactions (IIP, along the Y-axis). Amino acid residues are marked in single letter code. Jayashree et.al. Biology Direct 14 (2019)

As mentioned earlier, water fills about 30% of the interface of an average transient protein-protein complex. Water molecules form H bonds with protein groups of polar and charged side chains and can mediate polar interactions within the interface. Some of the waters associated with the free form are of course lost when the proteins bind indicating their role in stabilizing the structure of the free form of the protein. Those that remain in the interface of interacting proteins mediate inter-protein hydrogen bonds and should not be overlooked when analyzing possible key amino acids involved in the forces holding the proteins together.

Predicting Protein-Protein Interactions and Databases

The physical interactions of proteins in cells or purified are important for the identification and investigation of how proteins interact. Experimental methods to screen and identify interactions are costly and time-consuming. While helpful, computational docking predictors use structural and chemical complimentary scoring to find a possible binding site. These are not always accurate but do allow a targeted bench/wet lab approach to interrogate the interaction. When one adds structural changes that occur with or without substrate (take for instance the open-closed conformation of citrate synthase upon substrate binding) these models are only predictive. The structural database of complexed proteins is rapidly growing and coupled with the ability to accurately predict the 3-D structure of proteins has created a wave of computational methods to predict which proteins interact and where these interfaces might be located. These methods can be roughly divided into several categories: Protein-protein docking and modeling software based on existing single protein structures, and computational approaches using machine learning and neural networks focused on structure and sequence approaches. The former approach includes programs like Molecular Operating Environment (MOE) and Hawk Dock supports the prediction of key residues and sites of interaction of proteins. These predictive modeling programs are supported with cross-linking data to limit the possible interaction sites. There are several databases that collect and identify known and predicted interacting proteins. The numbers of protein complexes in these sites and the numbers of databases is growing. Several well-known major databases are: Data Visualization for Protein-Protein Interactions [GPS-Prot], A wiki based search engine based from Pitt [Wiki-Pi], A genomic collection of text mining and high-throughput experiments [String], the Biological General Repository for Interaction Datasets [BioGRID], the Molecular INTeraction database [MINT], Hypothetical and possible interactions from with links from publications turned into a conceptual network [iHOP], the IntAct molecular interaction database [IntAct] and the Human Protein Reference Database [HPRD]). Each database will contain information with different approaches of predicted and/or published interactions. Interestingly interactions identified in one database may not be found in others.

Suggested Readings:

- Srere, P. A. The metabolon. *Trends Biochem. Sci.* **10**, 109–110 (1985)
- Zhang et. al. Protein-Protein interactions and metabolite channeling in the plant tricarboxylic acid cycle. Nature Communications 8, 15212 (2017)
- Zhang and Fernie. Metabolons, enzyme-enzyme assemblies that mediate substrate channeling, and their roles in plant metabolism. Plant Commun. 2(1) eCollection (2021).
- Kuznetsova et. al. What Macromolecular Crowding Can Do to a Protein Int J Mol Sci Dec; 15(12): 23090–23140. Published online 2014 Dec 12. (2014)
- Scheiber. Kinetic studies of protein-protein interactions. Current Opinion in Structural Biology 12(1) pp 41-47 (2002)
- G. Schreiber, CHAPTER 1:Protein–Protein Interaction Interfaces and their Functional Implications , in Protein–Protein Interaction Regulators, 2020, pp. 1-24
- De Las Rivas and Fontanillo Protein-Protein interactions essentials: key concepts to building and analyzing interactome networks. PLoS Comput Bio 6 (2010)
- Ozbabacan et. al. Transient protein-protein interactions. Protein Engineering, Design & Selection 24(9) p 635-648) (2011)
- Perkins et. al. Transient Protein-Protein Interactions: Structural, Functional and Network Properties. Structure 18(10) p 1233-1243 (2010)
- Srinivasan et. al. Interface residues of transient protein-protein complexes have extensive intra-protein interactions apart from inter-protein interactions. Biology direct 14,(1) (2019)