

Synopsis

The functioning of all living systems, from prokaryotes to multi-cellular, higher-order organisms, is orchestrated by precise physico-chemical principles that embody macromolecular interactions. Molecular recognition constitutes the basis of these macromolecular associations. It involves biological macromolecules interacting with various small molecule ligands with optimal specificity and affinity to form a specialized functional complex. Biological ligands constitute a diverse repertoire of entities ranging from light, ions, peptides, proteins, hormones, drugs and neurotransmitters. The association between a protein and its cognate ligand results into a cascade of molecular re-organizations that eventually form a functional protein-ligand complex poised to perform a precise biological role. Biophysical techniques provide information about binding kinetics and associated thermodynamics encompassing the binding event. Computational perspectives on mining protein-ligand interactions involve docking, binding free energy calculations and drug-design. In spite of a plethora of available techniques, the charisma underlying the specificity and subtlety of bio-molecular associations remains unexplored and hence continues to attract researchers.

The studies presented in this thesis address the conformational plasticity inherent to proteins by virtue of which they are able to adapt to diverse ligands and orchestrate complex biological processes. To this end, we analyse proteins as mathematical formulations and apply these formulations to better comprehend protein-ligand interactions. Broadly, the objective of the thesis is to mechanistically understand protein-ligand associations at a molecular detail.

We investigate several proteins and their protein-ligand complexes from network formalism for transforming a protein three-dimensional structure as a network comprising of nodes and its connections. In this process, we study the mechanistic aspects of ligand binding by analysing the response of such Protein Structure Networks (PSNs). Degree of similarity in PSNs of ligand-bound complexes has been quantified by means of a metric called Network Similarity Score (NSS) and its applications have been examined. We have also probed the physical nature of such networks by subjecting them to targeted 'perturbations. Effects of such local disturbances have been carried out by either deleting a component node or its associated connections in the network; and its effects have been analysed. Major players in maintaining the stability of such networks have also been identified. Another interesting detail captured by network formalism is residue-wise grouping or clustering. This concept utilises graph spectral methods and yields insights about the ordering of residues (at backbone and side chain levels) in a PSN. The biological applications of these concepts have been extensively explored in this thesis.

The Introduction (Chapter 1) outlines the concepts underlying protein-ligand interactions and provides a summary of experimental and computational means of understanding them. The methodologies adopted have been outlined in Chapter 2. These Chapter 4 presents extended applications of PSNs towards understanding protein-protein interactions. With a view to understand this phenomenon, the classical GPCR-G-protein complex was revisited with an aim to understand agonist- and antagonist-induced changes at the protein-protein interface. These protein-protein associations were inspected in terms of weighted side-chain edges and bring to light distinct clustering patterns that delineated ligand-induced conformational changes. A weighted networks methodology was adopted in order to understand the propagation of information across the 7TM architecture in terms of non-covalent side-chain

interactions. Analysing binding site architecture, side-chain contacts and finally residue level clustering provided details about the ability of the proteins to relay signals in the presence of agonists and associate with G-proteins.

The subsequent two chapters discuss the effects of protein-ligand associations in a dynamical perspective. The behaviour of proteins as dynamic entities was captured by observing them through Molecular Dynamics (MD) Simulations. Trajectories of proteins reveal the diverse range of adjustments that a protein makes upon binding to a ligand. Analysis of MD trajectories provides a powerful tool to explore the conformational space sampled by proteins *in silico*. To this end, all-atom simulations for members of a ligand-inducible transcription factor superfamily, i.e., Nuclear Receptors have been carried out. In particular, two members of this family have been studied, namely the Pregnane X Receptor (PXR) and the Estragon Receptor α (ER α). The examples portrayed extremes of ligand binding behaviour, wherein PXR binds a variety of ligands and is classified as a promiscuous' receptor and ER α is a highly specific in its action since it binds to the hormone estrogen.

Chapter 5 discusses the ligand induced effects on PXR and its binding partners i.e. Retinoic acid receptor (RXR) and associated transcriptional coactivators. Earlier studies in the lab reported the presence of an accessory, druggable ligand binding pocket that was observed exclusively through simulations (Chandran and Vishveshwara, 2016). The dynamics of this pocket were explored in response to PXR's association with RXR and coactivators. Principal component analysis, pocket volume calculations and calculation of statistical free energy revealed changes in behaviour of this accessory binding pocket in response to complex formation with RXR and binding to ligands. Further, the role of heterodimerization was found to be important in maintaining the balance between promiscuity and specificity of these complexes. Chapter 6 involves an in-depth molecular dynamics study on Estrogen Receptor α (ER α), where the effects of agonist, antagonist and endogenous ligands were studied at various scales. Our analysis provided details about the underlying specificity of this receptor in comparison with the promiscuous nature of PXR.

The last section of this thesis outlines a study on similarity of binding sites and in diverse proteins. It explores the percolation of binding site similarity to regions beyond the binding site. Chapter 7 discusses how similarity at binding site advances to regions beyond them. An in-house algorithm, Pocket Match was used to elucidate this expanded similarity in binding sites and other regions. Our results illustrate this extended similarity and are discussed with case studies and examples.

A brief summary of the results and their future directions are outlined in the final chapter of this thesis (Chapter 8). The application of network formalism to investigate several problems of biological relevance has been outlined in the preceding chapters. One can apply network based studies to predict mutation sites and other biophysical studies may be performed to further complement our computational outcomes

