Abstract

Proteins are biological macromolecules made up of linear chains of amino acids, and are organized into three-dimensional structures comprising of different secondary structural ele- ments. In its functional form a protein acts like a complex network where the nodes are the constituent amino acids, and the links are the chemical interactions that hold them together with short and long-range contacts. Thus protein three-dimensional structures can be mod- elled using Graph Theory as complex networks of interacting amino acids. These are termed Protein Contact Networks (PCN). Since many topological properties of networks can be understood from the network parameters, we believe that it can also be a useful approach to identify the different structural classes of proteins and their in uence in protein function. In this study, we have attempted to understand how network properties and attributes can be used to study - (a) the major structural classes of proteins, and (b) the relationship between structure and function in proteins, which do not show significant conformational changes in ligand-binding. As per the Structural Classification of Proteins (SCOP), proteins are grouped into four classes, i.e. α, β, α + β, and α/β based on their major secondary structural contents, which have different topologies. PCNs were developed for each class (50 proteins in each class) and different visual methods used to understand the differences among them. Several network parameters were calculated at local and global level, and their distribution studied. Average clustering coeffcients showed statistically significant differences among the classes, except between α + β, and α/β. The average shortest path did not show any difference among any class. The degree distribution and the number of residues having the most common degree show variation among the structural classes. Additionally, all PCNs of proteins of all classes showed small world nature. In our attempt to study structure-function relationship using the PCN approach, we used the HIV-1 Reverse Transcriptase protein (apo) and its ligand-bound form (holo) as the model systems. We used three structures - 1RTJ (HIV-1 RT unbound), 1IKW (HIV-1 RT bound to EFZ, which is a non-nucleoside reverse transcriptase inhibitor), and 1FKO (HIV-1 RT with resistance mutation at K103N). We calculated the root mean square deviation (RMSD) among the three structures, which was found to differ very little. This was corroborated by insignificant variations in the global clustering coeficient and average shortest paths of the three PCN. We then used the PCNs to study the loss and gain of contacts among the three networks with different functionality. We analysed the contacts in the ligand-binding pocket and interface between the two chains, and identified few important contacts that allow the change in function in spite of the three dimensional structure being quite similar. Thus, the work presented in this thesis argues that the complex network approach to study protein three-dimensional structure can not only be an important and useful methodology to study structural attributes of a protein, but can also unravel local changes in contacts for understanding protein structure-function relationship.