Chapter 6

Conclusions

After the synthesis in the cell, folding of the amino acid chain is important for attaining the structure required to reach a functional state as soon as possible. This happens through the formation of short- as well as long-range interactions. While the former are largely responsible for formation of secondary structure units, the latter bring spatially distant (along the chain) residues closer. Secondary and tertiary structures are formed primarily by noncovalent interactions. Our graph theoretical representations of proteins structure, Proteins Contact Network (PCN) and Long-range Interaction Network (LIN), model various aspects of the three-dimensional structure of a protein in an attempt to understand its function and kinetics.

The Small World Nature

We found that proteins of diverse structural and functional classification have small-world nature with low characteristic path length \(L\) and high level of clustering \(C\) as shown in Figure 6.1. In this regard PCNs are similar to most other real-world networks. Interestingly, we find that LINs depart from the small-world nature. The LINs have medium range of \(C\) in the proteins studied. The implication of small-world nature of PCNs is attributed to the case of dissipation of energy upon complexation. Such a property may have important role in efficient allosteric regulation of protein functions.
We find that PCNs are characterised by hierarchical nature, as shown by the independence of their clustering coefficient with size (Figure 3.10). This observation is in accord with earlier findings by other researchers \[4, 74, 75\], and more investigation needs to be done on this line. We would like to point out that it has been found in many real networks \[32\] the hierarchy and modular architecture go hand in hand. Our preliminary results suggest modular architecture in PCNs. It would be interesting to see what significance, if any, the modules thus found in protein structures would have. Such modules could be identified by the network community structure algorithms.

**Assortative Nature of PCNs and LINs**

In contrast to all other naturally evolved intracellular networks studied so far, we found that contact networks of proteins show assortative mixing at both short and long length scales i.e. rich nodes tend to connect to other rich nodes. This is an exceptional property as all other real-world networks known (except for social networks) are disassortative. Interestingly, we find that LINs too are assortative, which implies that assortativity is independent of short-range interactions. We built appropriate random controls to identify
the appropriate network feature that possibly contributes towards assortativity. We found that degree distribution contributes significantly towards assortative mixing in PCNs as well as their LINs.

The predominance of disassortativity in real-world networks have been alluded to confer the property of robustness (reduced spread of perturbations) in the network. Then why are the contact networks of protein structures assortative? Communication among the residues of the protein is important. It is known that “network of residues” mediate allosteric communication in proteins [72, 73]. It is also proposed that allostery is an intrinsic property of all dynamic proteins [129]. We propose that assortativity is an indicator of ‘allostERIC communication network’ established within the protein structure and is important enough to be found in all proteins.

The role of specific residues in protein folding and their evolutionary conservation is highly debated [64, 130, 131]. Mirny et al. [64] found that conserved residues that are part of folding nucleus, across proteins, were found to be in contact with each other. Based on this finding, we propose that folding nucleus of a protein could be a subset of the set of residues that form assortative group.

Our observation of assortative mixing and hence a set of residues that are part of a assortative network opens up new directions of work.

Biophysical implication of topological parameters

One would expect to have biophysical implications of the exceptional network properties that we observe. We found that for both PCNs and LINs, coefficient of assortativity, a measure of the assortativity, has positive correlation with the rate of folding of single-domain, two-state folding proteins. Similarly, we find that clustering coefficient of LINs has a high negative correlation to the rate of folding of these proteins, though that of PCNs show no significant correlation. Other workers have developed parameters specific for proteins (CO, LRO, TCD) and correlated with rate of folding. Our aim was to show the relevance of general network parameters to a kinetic property of the proteins. Indices such as closeness, betweenness offer more local and hence residue-specific information. By combining our general, global parameters with such local ones one could address broader questions related to
protein structure and function.

Advantages and Limitations of PCN Model

PCN is, by virtue of coarse-graining, a simple model. It doesn't involve the time evolution of the protein structure. Rather it models the static native state structure. We don't explicitly consider information about the chemical nature of the side-chains in this model. Although, since the final native-state structure is an outcome of the chemical interactions happening among various amino acids the model implicitly does consider the chemical interactions involved.

Each of the twenty amino acids has different numbers of atoms and hence has different size. Hence the nature of noncovalent contacts depends on the specific amino acids involved. We haven't included that information in our model so far.

The time evolution of the protein structure can be considered by building the weighted network using the Transition State Ensemble (TSE) structures. Depending on the question being asked and its sensitivity to the above-mentioned details one may consider adding further details to the PCN, thereby enhancing it.

Thus complex network analyses offers to be an important tool in studying the structure-function of proteins—the fascinating molecule of life.