# Mutations in protein family networks

**Nivedita Deo**
*University Of Delhi, Delhi 110007, India, Delhi, India*

To study the structural organization of a protein family we develop an approach based on the random matrix theory (RMT) and Network theory, which uses the physiochemical properties of the amino acid with multiple sequence alignment. A novel graphical technique is designed to represent protein sequences using physiochemical properties that give a fast, easy, and informative way of comparing the evolutionary distances between protein sequences. The first step is to calculate the correlation matrix associated with each property, where the noise reduction and information filtering are done employing RMT using an ensemble of Wishart matrices. The investigation of the eigenvalue statistics of the correlation matrix for the beta-lactamase family shows the universal features as seen in the Gaussian orthogonal ensemble (GOE). The statistical and spectral analysis of the Pearson correlation matrices between positions based on physiochemical properties of amino acids of several protein families is performed and compared with the random Wishart matrix model results. A detailed analysis shows that the protein families significantly diverge from the Marcenko-Pastur distribution with many eigenvalues (outliers) outside the Wishart lower and upper bound. The information content of each eigenvector of the correlation matrix is quantified by introducing an entropic estimate, which shows that for the smallest eigenvectors (low eigenmodes) are highly localized as well as informative. These small eigenvectors when processed give clusters involving positions that may have well-defined biological and structural importance.

Next, we use the correlation-based threshold method to create a weighted multiplex network of evolutionary interactions between positions for the β-lactamase family. Although each network layer is derived from the same multiple sequence alignment (MSA) but is diverse from other layers by realizing a different aspect of the interaction between amino acids. In each layer, positions show a different clustering, where the clusters depend on the physiochemical properties, and the positions with similar strength of the property tend to interact with each other. This clustering gives a neighborhood of highly interconnected nodes with strong interaction, indicating a high possibility of functional and evolutionary constraints, sometimes the part of the sector that forms the important functional and structural domain. We observe a hierarchy in the physiochemical properties, some interactions in some layers are preferred over others, which can identify the property responsible for the functionality of the family. Multi degree also sets a hierarchy in the influence of properties at a given position. Patterns and regularities in the protein multiplex network are explored.
The effect of mutation on a protein family is analyzed by fixing a particular amino acid at a given position of the Multiple Sequence Alignment (MSA), and taking all the sequences which have the same amino acid ‘X’ at that position. Multiple subsets are created with a different types of amino acid at a fixed position. We then examine and compare these subsets as this gives the effect of a particular amino acid at that position. Preliminary results on perturbating the correlation matrices and networks are discussed.