# Functional motifs in protein families

P. Bhadola, N. Deo
University of Delhi, India

The structural organization of a protein family is investigated by devising a method based on the random matrix theory (RMT), which uses the physio chemical properties of the amino acid with multiple sequence alignment. A graphical method to represent protein sequences using physio chemical properties is devised that gives a fast, easy, and informative way of comparing the evolutionary distances between protein sequences. A correlation matrix associated with each property is calculated, where the noise reduction and information filtering is done using RMT involving an ensemble of Wishart matrices. The analysis of the eigenvalue statistics of the correlation matrix for the $\beta$-lactamase family shows the universal features as observed in the Gaussian orthogonal ensemble (GOE). The propertybased approach captures the short as well as the long range correlation (approximately following GOE) between the eigenvalues, whereas the previous approach (treating amino acids as characters) gives the usual short-range correlations, while the long-range correlations are the same as that of an uncorrelated series. The distribution of the eigenvector components for the eigenvalues outside the bulk (RMT bound) deviates significantly from RMT observations and contains important information about the system. The information content of each eigenvector of the correlation matrix is quantified by introducing the entropy of eigenvector components, which shows that for the $\beta$ -lactamase family the smallest eigenvectors (low eigenmodes) are highly localized as well as informative. These eigenvectors corresponding to the smallest eigenvalue when processed gives clusters involving positions that have well-defined biological and structural importance matching with experiments. The approach is crucial for the recognition of structural motifs as shown in $\beta$-lactamase (serine protease, HSP70, G protein, HTH 1 and other families) and selectively identifies the important positions for targets to deactivate (activate) the enzymatic actions.

[l] P. Bhadola, N. Deo, Phys. Rev. E **94**, 042409, (2016).
[2] N. Halabi, O. Rivoire, S. Leibler, R. Ranganathan, Cell **138**, 774 (2009).
[3] S. Cocco, R. Monasson, M. Weight, Plos Compt. Biol. **9**, e1003176 (2013).
[4] O. Rivoire, Phys. Rev. Lett. **110**, 178102 (2013).