

## Where informatics lags chemistry leads

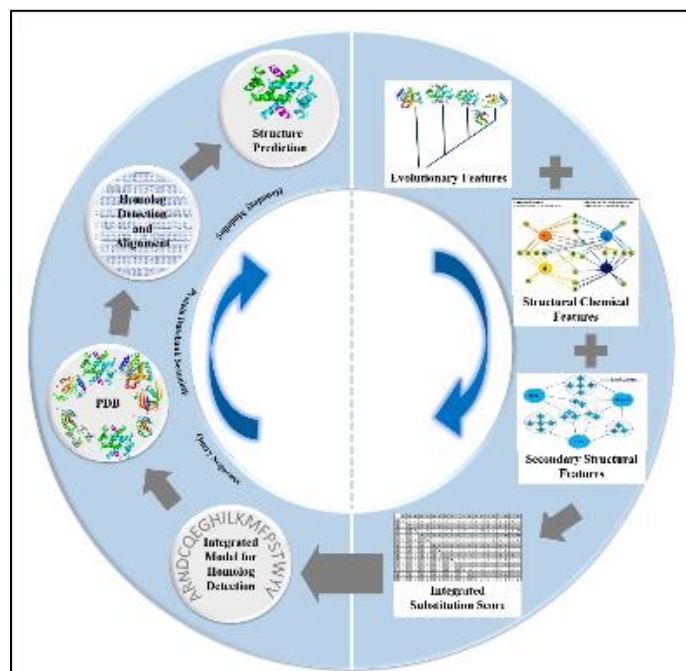
Rahul Kaushik<sup>§, #</sup>, Ankita Singh<sup>#, €</sup>, B. Jayaram<sup>§, #, ¥, \*</sup>

§ - Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, India

# - Supercomputing Facility for Bioinformatics & Computational Biology, Indian Institute of Technology Delhi, India

€ - Department of Bioinformatics, Banasthali Vidyapith, Banasthali, India

¥ - Department of Chemistry, Indian Institute of Technology Delhi, India



## **ABSTRACT**

That amino acid sequences dictate the tertiary structures of proteins has been known for more than five decades. While the molecular pathways to tertiary structure are still being worked out, with the axiom that similar sequences adopt similar structures, computational methods are being developed continually in parallel, utilizing the protein databank structural repository and homolog detection strategies to predict structures of sequences of interest. The success of this approach is limited by the ability to unravel the hidden similarities among amino acid sequences. We consider here the 20 amino acids as a complete set of chemical templates in the physico-chemical space of proteins and propose a new structural and chemical classification of amino acids. An integration of this perspective into the conventional evolutionary methods of similarity detection, leads to an unprecedented increase in the accuracies in homolog detection further resulting in improved protein structure prediction. The performance is validated on a large dataset of 11716 unique proteins and the results benchmarked against conventional methods. Availability of good quality protein structures helps in structure based drug design endeavors and in establishing protein structure function correlations.