

Postgraduate student: **Mr. Anuj Sharma**

Thesis Title:

Improving sequence similarity search for Protein Homology Inference using Structural data and Machine Learning methods

Abstract:

In this thesis, we presented a new work flow for improving Homology detection for Proteins. The proposed methodology combines protein sequence and structure comparison data for detecting Homology. Sequence similarity measures are the most commonly used tool for homology detection. However it is known that evolutionary divergence can lead to homologous proteins having very little sequence similarity. In the middle range of proteins found in the sequence similarity results this divergence leads to errors in homology classification. This zone of proteins where maximum error occurs is referred to as the "twilight zone" of proteins. The work flow presented involves reclassifying twilight zone proteins, in the PSI-BLAST results, into 'true positives' and 'true negatives'. The reclassification is performed using classifiers built from the structural data.

Several parametric, non-parametric and committee classifiers have been compared on standard metrics. Classifiers were built using structure comparison data and subsequently used for reclassifying the twilight zone proteins. We provide statistical data supporting the separability of the two classes (homologous vs. non-homologous proteins) and subsequently provide results of classification using various classifiers tested. Various combinations of structural features extracted were tried. Our results show that the proposed sequence-structure combined work flow can improve a homology detection by reducing significantly errors that occur in the 'twilight zone' when plain sequence comparison is used.

Examining Committee

Elias Manolakos, Assoc. Professor, Department of Informatics and Telecommunications, University of Athens

Yiannis Emiris, Professor, Department of Informatics and Telecommunications, University of Athens

Dr. Evangelia Chrysina, Dr. Evangelia D. Chrysina Structural Biology & Chemistry Group
Institute of Organic & Pharmaceutical Chemistry, National Hellenic Research Foundation