Rushi Verma (2010). *Annotation of Plasmodium Falciparum Proteins using Machine Learning Techniques and Evolutionary Information.* Ph.D. Thesis. CSIR-IMTECH, Chandigarh/HPU, Shimla.
Co- Supervisor: Dr. GPS Raghava
Th-227

SUMMARY OF THE THESIS

Forty-one percent of the world's population live in areas where malaria is transmitted (e.g., parts of Africa, Asia, the Middle East, Central and South America, Hispaniola, and Oceania). In 2002, malaria was the fourth cause of death in children in developing countries, after perinatal conditions (conditions occurring around the time of birth), lower respiratory infections (pneumonias), and diarrheal diseases. Malaria caused 10.7% of all children's deaths in developing countries. Wrongly, the incidence of malaria is presently rising due to the emergence of drug resistant strains of the malaria parasite and population growth of increasing insecticide-resistant mosquito vector. The life cycle of *Plasmodium* passes in two hosts, a vertebrate and a mosquito. Propagation in the vertebrate host is initiated by the bite of the infected female *Anopheles* mosquito, which introduces sporozoites in to the blood stream. The effect of the disease is associated with the intraerythrocytic stage where parasite develops through ring (0-24 hrs), trophozoite (24-36hrs) and schizoint (36-48hrs) finally lysing the host cell for merozoite liberation and subsequent invasion to newer RBCs. The parasitism of erythrocytes by malaria organisms significantly alters the physiological functioning and cellular biology of these enucleated host cells, resulting in conditions, in favor of its survival and growth. Several among the membrane-associated molecules have been projected as prospective candidates for vaccine development. Various studies have also demonstrated a significance presence of a motif which helps in the export of the parasite proteins to the infected RBC and further upto the iRBC's surface, which can prove to be a potential drug targets. Also they may be some physio-chemical properties or structural differences which may govern the localization of the parasitic proteins.

The main objective of this research work was to develop bioinformatics tools for identification of sub cellular localization of malaria parasitic proteins which can be crucial for vaccine design. The steps include first the identification of the secreted proteins which crosses the PVM and beyond in the absence of any signal or motif. In this we were able to archive a good accuracy by multi model system. Accordingly, our first model is motif-based that uses MEME/MAST for predicting those proteins which have PEXEL/VTS signal motifs. Secondly, in order to predict PF erythrocyte membrane protein (PfEMP) type proteins, we developed a domain-based model using HMMER. Thirdly, a SVM based model was developed, using composition of proteins, for predicting those PF secretory proteins which neither have motif nor domain. This SVM model was trained and tested on experimentally validated secretory proteins. Maximum MCC 0.58, 0.57 and 0.68 were achieved using amino acid, dipeptide and PSSM composition respectively. Secondly we distinguish the mitochondrial targeted proteins of the PF as they are also a part of the secreted proteins and compared their genome with other mitochondrial proteins of various species. A SVM model for distinguishing mitochondrial and non-mitochondrial has been developed using composition of PSSM profile which achieved MCC 0.75 and accuracy 91.38%.We then short listed some PF proteins which can be potential vaccine candidates using bioinformatics approaches .