**Sneh, Lata (2009).** *Bioinformatics approach for searching subunit vaccine candidates and adjuvants.* Ph.D. Thesis. CSIR-IMTECH, Chandigarh/ Jawaharlal Nehru University, New Delhi: India.
**Supervisor: Dr. GPS Raghava**
Acc. No. : Th-203

## SUMMARY OF THE THESIS

Subunit vaccines have emerged as an important strategy for vaccine designing in an attempt to achieve strong immunogenicity without increasing 'reactogenicity'. Subunit vaccines contain only immunogenic regions of antigen instead of a complete antigen and are able to evoke an immune response often with fewer side effects than might be caused by a vaccine made from the whole organism. But many a times these subunit vaccines (antigenic peptides) show poor immunogenicity when tested experimentally, owing to the poor response generated against them. Ironically, adding back components from microbes which were previously removed from these antigens helps to enhance the efficacy of subunit vaccines. These components known as PAMPS are evolutionarily conserved signature pattern from microbes and enhance the efficacy of vaccines when mixed with it, thereby acting as adjuvants. It is now well established that for a subunit vaccine to work we need a cocktail of antigenic peptide and adjuvant.

PAMPs as potent adjuvants can improve the effectiveness of vaccines by accelerating the generation of robust immune response and reducing the amount of antigen needed, thus reducing the cost of vaccination program. Moreover, the expression of cytokines and chemokines is also necessary to build a response that is strong enough to combat the invading infection. Thus in this research work we have tried to focus on all these aspects so that we could develop methods that would help the immunologists to develop not just vaccine but effective vaccine candidates. We have also made an attempt to develop a method that could help the biologists to design peptides that would be antibacterial in action and can prove to be a solution to growing antibiotic resistance among the bacteria. We, therefore, planned to adopt a systematic approach to combine both arms of immune system (the two important steps of vaccine designing) i.e. the adaptive immune system and innate immune system, in order to get an optimized the immune response against a pathogen. Bioinformatics techniques were applied to focus on the selection of the identification of key innate immune targets for induction of potent, but safe, immune responses.

1) The first step towards an effective subunit vaccine design strategy is to discover and identify antigenic peptides that could be utilized as potential vaccine candidates. In this direction we developed prediction methods that could not only qualitatively but also quantitatively determine the binding capabilities of a peptide for a MHC.

We developed a method HLA_Affi for predicting affinity of peptides that bind to HLA-A2 allele. We used various input patterns like binary pattern, amino acid composition, dipeptide composition, physico-chemical property composition and various combinations of these to train the method. A maximum correlation of 0.59 (between the actual and predicted binding affinity) was achieved by SVM model that was trained on Binary pattern + physico-chemical property composition hybrid pattern. The method is available on http://www.imtech.res.in/raghava/hla_affi/.

We also developed a method FDR4 to predict the affinity of peptides that bind to a MHC class II allele, HLA_DRB1*0401. As the prediction of MHC class II binders is difficult due to their variable length, Support Vector Machine based optimization technique (SVMOT) has been developed to detect binding core in the known HLA-DRB1*0401 binders. Successive application of SVMOT was done to improve the accuracy of core prediction. The method SVMOT has been optimized in three cycles and performance was evaluated after each cycle on an independent dataset. A maximum correlation of 0.70, 0.82 and 0.87 was achieved between predicted and actual affinity of HLA-DRB1*0401 binders for first, second and third cycle of SVMOT using composition based strategy. Similarly, binary-pattern based strategy gave a maximum correlation of 0.67, 0.78, 0.79 between the actual and predicted affinity values of HLA-DRB1*0401 binders. Based on above approaches, a user-friendly web server FDR4 has been developed, which allows users to predict HLA-DRB1*0401 binders and their affinity in their protein sequence (http://www.imtech.res.in/raghava/fdr4/ and mirror site http://bioinformatics.uams.edu/raghava/fdr4/).

2) We also collected and compiled the much scattered information about the PRRs and PAMPS (that could be used as potential adjuvant candidates) in a database PRRDB. The current version of database contains around 500 patterns recognizing receptors from 77 distinct organisms ranging from insects to human. This includes 177 Toll-like receptors, 124 are Scavenger receptors and 67 are Nucleotide Binding Site-Leucine repeats rich receptors. The database also provides information about 266 ligands that includes carbohydrates, proteins, nucleic acids, glycolipids, glycoproteins, and lipopeptides. A number of web tools have been integrated in PRRDB in order to provide following services: i) searching on any field; ii) database browsing; and iii) BLAST search against the pattern-recognition receptors.

PRRDB also provides external links to standard databases like Swiss-Prot and Pubmed. The collection and compilation of PRRs and PAMPs, structures of PAMPs would help the scientists to analyze their fine structural details and study what is the minimal region required for the desired activity which is also safe at the same time.

3) In the next step we collected the structure of PAMPs from PRRDB and did analysis of various TLR binders. We performed an analysis of various physico-chemical properties of like Molecular Mass, Volume, Surface area, No. of H-bond donors/acceptors etc. The analysis of the physic-chemical properties of various PRR ligands could help to derive general rules that determine the ability of a molecule to activate PRRs. However, we do feel that the analysis performed above is done with a small number of ligands (that may to be just a small subset of PRR ligand when many more ligands will be discovered in future) and may not be the actual representation of the PRR ligands.

4) Next we tried to develop prediction methods that could predict cytokines, chemokines and chemokine receptors. We developed a method CytoPred, which is a PSI-BLAST + Support Vector Machine-based hybrid approach that could predict as well as classify cytokines. CytoPred is capable of predicting cytokines with an accuracy of 98.29%. The overall accuracy of classification of cytokines into four families and further classification into seven subfamilies is 99.77 and 97.24%, respectively. A user-friendly server CytoPred has been developed and available at http://www.imtech.res.in/raghava/cytopred.

We developed another method, ChemoPred, which is capable of predicting chemokines and chemokine receptors with an accuracy of 95.08% and 92.19% respectively. The overall accuracy of classification of chemokines into three subfamilies was 96.00% and that of chemokine receptors into three families was 92.87%. The server chemopred is freely available at http://www.imtech.res.in/raghava/chemopred/.

4) In the end we developed a prediction method that predicts potential antibacterial peptides. The dataset for development of prediction method was fetched from the database APD. First, the N-terminal residues were used for predicting antibacterial peptides using Artificial Neural Network (ANN), Quantitative Matrices (QM) and Support Vector Machine (SVM), which resulted in an accuracy of 83.63%, 84.78% and 87.85%, respectively. Then, the C-terminal

residues were used for developing prediction methods, which resulted in an accuracy of 77.34%, 82.03% and 85.16% using ANN, QM and SVM, respectively. Finally, ANN, QM and SVM models were developed using N and C terminal residues, which achieved an accuracy of 88.17%, 90.37% and 92.11%, respectively. All the models developed in this study were evaluated using five-fold cross validation technique. These models were also tested on an independent or blind dataset. In the later part of the chapter we also describe the improved version of antibacterial peptide prediction method, AntiBP2. AntiBP2 is a SVM based prediction method which is build on a larger dataset (almost double the size of the dataset used in AntiBP) of antibacterial peptide.

## 9.2 Future Prospects

Use of bioinformatics approaches would certainly improve the predictive capacity for scientists working on effective vaccine development. The collection and compilation of PRRs and PAMPs, structures of PAMPs would help the scientists to analyze their fine structural details. It would help them to look for the minimal region required for the desired activity which is also safe at the same time. The analysis of the adjuvants would certainly help the immunologists to identify the trend followed by these PAMPs that could act as potential adjuvants. All the databases and methods developed in this thesis will provide an illuminated way to immunologist for designing effective subunit vaccines against deadly diseases. These methods can help to obtain the relevant immunological information from unmanageable amount of data. In summary, these methods will act as stepping-stones in *in silico* effective subunit vaccine designing.