Chapter 5 Conclusion and future work

A proteomics research of plant diseases requires analysis of large sequences and is heavily dependent upon bioinformatics computation. Computational modeling should be performed according to strict standards, requiring careful data selection for protein model building, followed by adequate testing and validation through bioinformatics tools. The computational integration of such data is proving to be the most effective route to determine the protein function very quick and accurately. A database can provide easy access to previous results from the experiments and literature surveys, preventing the wasteful duplication of research. A well-designed database also supports both expert and machine guided searches for novel correlation in data. Integration of large datasets and information reveals from same resources in the life sciences is one of the most challenging goals facing biotechnology and pharmaceutical industries today. Software for prediction for ORF, binding sites, and gene helps molecular biologist make sense out of unmapped DNAs. The database component of the platform includes general information on medicinal plants, plant viruses, and sequence analysis with focused emphasis on proteomics, genomics, tissue culture, metabolomics, ethno-botanical, patent and related literature on the plants. Virus involved multiple antigenic components to direct and empower the immune system to protect the host from infection. MHC molecules are cell surface proteins, which take active part in host immune reactions and involvement of MHC class in response to almost all antigens and it give effects on specific sites. Predicted MHC binding regions acts like red flags for antigen specific and generate immune response against the parent antigen. So a small fragment of antigen can induce immune response against whole antigen. This theme is implemented in designing subunit and synthetic peptide vaccines. The sequence analysis method is allows potential drug targets to identify active sites, which form antibodies against or plant diseases. The method integrates prediction of peptide MHC class binding; proteosomal C terminal cleavage and TAP transport efficiency. Antigenic epitopes of capsid protein are important antigenic determinants against the various toxic reactions and viral infections.

The system also provides a comprehensive selection of data analysis and simulation tools alongside an advanced query system and a context-mapping tool that

implements a relevancy model towards correlating various data sources. It is need of new genomic era, to reduce the laboratory work, time, chemicals and analysis of complex genetic algorithms to solve the biotechnological problems. This will help in development of new drug discovery and development process.

A proteomics research of plant diseases requires analysis of large sequences and is heavily dependent upon bioinformatics computation. Computational modeling should be performed according to strict standards, requiring careful data selection for protein model building, followed by adequate testing and validation through bioinformatics tools. The computational integration of such data is proving to be the most effective route to determine the protein function very quick and accurately. A database can provide easy access to previous results from the experiments and literature surveys, preventing the wasteful duplication of research. A well-designed database also supports both expert and machine guided searches for novel correlation in data. Integration of large datasets and information reveals from same resources in the life sciences is one of the most challenging goals facing biotechnology and pharmaceutical industries today.

Specific contribution of this research

- Plant virus database provides a central source of information about 1837 viruses, viroids and satellites of plants, fungi and protozoa, with some additional data on related viruses.

- This database prepared describing Accession Number of Plant Virus, GenBank Identification Number of Proteins from Virus, Taxonomy ID, Species Name, Protein Name, Taxonomy (Classification of Plant Virus), FASTA of coat Protein of Plant Virus, Reference for protein sequences and virus data.

- Gene Tool is helpful for those scientists who want a quick result before research and it also helpful for ORF prediction in genetic engineering and molecular biology. The user can compare the original sequence and the input sequence.

- This work provides insight into the molecular history for the gene families of the viral species, linking the expansion of these families to the duplications of

the gene cluster regions, and showing that they are composed of subgroups with distinct evolutionary (and possibly functional) differences.

- The evolutionary relationships among the potyviridae coat protein sequences were depicted, which shows changes in coat protein sequences. The challenges faced when different pesticides are not counteract the paste control; because of these mutations of same coat protein sequences get different results on plant to protect from the viral attack. Such regions represent conserved functional or structural domains in predicting the function and structure of proteins from plant viral species.

- Coat protein involved multiple antigenic components to direct and empower the immune system to protect the host from infection.

- MHC molecules are cell surface proteins, which take active part in host immune reactions and involvement of MHC class in response to almost all antigens and it give effects on specific sites.

- Predicted MHC binding regions acts like red flags for antigen specific and generate immune response against the parent antigen.

- Small fragment of antigen can induce immune response against whole antigen. This theme is implemented in designing subunit and synthetic peptide vaccines.

- The sequence analysis method is allows potential drug targets to identify active sites, which form antibodies against or viral infection.

- The method integrates prediction of peptide MHC class binding; proteosomal C terminal cleavage and TAP transport efficiency.

- Antigenic epitopes of viral coat protein are important antigenic determinants against the various toxic reactions and viral infections.

Limitation

The last several decades have witnessed a vast accumulation of biological data and data analysis. A more complete understanding of a biological system is gained when different types of data and conclusions drawn from the data are integrated into a larger-scale representation or model of the system. Ideally, this type of model is consistent with all available data about the system, and it is then used to generate

additional hypotheses to be tested. Computer-based methods intended to formulate models that integrate various events and to test the consistency of these models with respect to the laboratory-based observations on which they are based are potentially very useful. In addition, in contrast to informal models, the consistency of such formal computer-based models with laboratory data can be tested rigorously by methods of formal verification.

Future work

Computer science and mathematics are transforming molecular biology from an informational to a computational science. Drawing on computational, statistical, experimental, and technological methods, the new discipline of computational molecular biology is dramatically increasing the discovery of new technologies and tools for molecular biology.

Computational simulation will expand in prolific way, given a need to obtain a greater degree of predictability in animal and human clinical trials. Added to this, is interesting scope that '*in silico*' testing provides to deal with the brewing hostility towards animal testing.

Gene analysis work may be extended for the other database and reduce the number of missed minutiae by improving the performance for better performance of system. Also the work may extend to develop the new techniques and algorithms for fragment based subunit and synthetic peptide vaccines.

The data will be exported regularly in the advance database architecture, with incrementally performing major revisions, recoding and updating data-mining software. A data structure can be viewed as an interface between two functions or as an implementation of methods to access storage that is organized according to the associated data type. Data warehouse is created by analyzing ways to categorize data using dimensions and ways to summarize data using measures. Dimensions can be used to filter data by excluding results or by displaying data in different cells of a presentation. Measures are used to create averages and totals using precomputed aggregates.

Recently developed genomic technologies, such as oligonucleotide microarrays and achievements including whole genome sequencing, have suggested that scientists can now analyze complex genetic diseases at a much more rapid pace. The analytic speed is further increased by the large amount of genetic and genomic information that is

available in public databases, which enables several analytic steps to be computationally performed. This helps in development of new discovery and development process in life sciences.