12. Summary and future perspective

In this study, we have explored the various area related to computer-aided drug design based on protein-ligand interactions. The study has been divided into total 4 parts (i) Target/protein-based drug design; (ii) Ligand/peptide-based drug design; (iii) Designing of drug delivery vehicles; and (iv) Screening of drug targets.

In the Target/protein-based drug design, we tried to understand protein-ligand interactions by analyzing the various properties associated with the target/protein. This section was further subdivided into two parts (a) protein-small molecule interactions; and (b) protein-peptide interactions. In the first section, we created the number of non-redundant, high-quality datasets using PDB March 2018 release. Using the above data, we extracted binding site information of 901 ligands. As all ligands are not biologically relevant, we excluded irrelevant ligands from this dataset finally leading to 544 biologically valid ligands. Molecular properties of the ligands and as well as their binding site properties were further analyzed. Clustering analysis was also performed using the above features. Next, we made a comparison between sequence and structure-based methods developed for predicting ligand interacting residues. We found that the sequence-based method performs better in comparison to structure-based methods. In the last section of this study, an in silico method SAMbinder was developed for predicting SAM binding residue in a given protein sequence using evolutionary information as an input feature. In the second section, we benchmarked several molecular docking tools on various datasets. We performed blind and re-docking studies and observed that re-docking studies performed better in comparison to blind docking. We found that FRODOCK performed best in case of blind docking and ZDOCK performed better in case of re-docking. Web-based service, PPDbench has been developed where user can compute CAPRI parameters for its docked poses.

In the next part of the study i.e. second part, we developed computational tool Antifp, for predicting the antifungal nature of the natural peptides as well as a tool for predicting the antimicrobial nature of the chemically modified peptides. In the case of antifungal peptide prediction, we observed that our SVM based model performed best on the N15C15 amino acid composition. In total, three datasets were used in the study and model corresponding to each dataset has been incorporated into the webserver. In the next part, we analyzed the chemical modifications, which makes a peptide antimicrobial or non-antimicrobial. Several features (atom & diatom composition, 2D descriptors, various fingerprints, binary profiles using SMILES format) were computed, and machine learning models were developed.

Among these models, fingerprints based models performed best and were incorporated into a web server known as AntiMPmod.

The third section of the study deals with the information of peptide-based drug delivery vehicles. These peptides are popularly known as cell-penetrating peptides. We developed a database CPPsite 2.0, where we incorporated information of 1850 natural and chemically modified CPPs. Several search modules and analysis tools are present in the database for the user. In the last section of the study i.e. fourth part, we have developed a web-based service where user can annotate the mutation information present in its protein of interest by comparing it with the previously reported binding sites of different ligands. The web site will provide the knowledge of the ligands whose binding site has been disturbed by the mutation at that particular position in the protein. We mapped the mutant residue with the ligand-binding region by creating a pattern of window length 17. If the residues present in the pattern except central residue is matching, we say that the ligand might interact with the mutant residue. We also provided the list of ligands which can interact with the mutant residue based on its propensity present in the binding site computed using experimentally determined structures. The study will help in finding new drug targets which can be exploited further for therapeutic purposes.

Overall, the study done in this thesis addresses various aspects of the protein-ligand interactions and provides useful insights.

Future Perspective

Protein-ligand interaction possesses wide application in the field of drug discovery. Understanding of the mechanism helps in selecting and optimizing the correct therapeutic targets and lead molecules for desigining novel drugs. In the current study, an attempt has been made to understand the nature of the ligands and its binding site in the proteins. The insight gain from the study was used for developing new algorithms as well as annotating the mutations at genome scale. However, there are certains points which still needs to be addressed for developing better understanding of the protein-ligand interaction and can be a part of futuristic study. In the below section we are discussing such few points which will help in improving the study further.

In the present study, few selective features were used for analyzing ligand nature. However, more features such as hydrogen bond donor and acceptor counts, surface area of the ligand, its association and dissociation constant values can be further used for understanding the

nature and similarity among ligands. Also, in the present study, we do not considered the effect of post-translational modification of the protein in its ligand binding affinity, therefore, there is need to consider these and other similar features to have robust and comprehensive understanding of the mechanism.

Benchmarking study suggests that there is need to develop more sophisticated and general method which can predict the ligand interacting residue in protein sequence for number of ligands as it is not possible to develop specific methods for all ligands. During our docking studies, we observed that the methods were unable to rank best pose as the top pose. Hence there is need to develop a better scoring function which can utilize the force-field, empirical, and knowledge based properties and rank the poses more accurately. This will help in finding the near-native structures which could help in desigining better drug molecules. Lastly, there is need to explore the area of genomics and we can annotate the structural and function of those mutants at genome level. With the advancement in the sequencing technology, we have number of sequence which needs to be annotated. Although there are few methods which have been developed addressing these problems; however, they are very generalized and provides information regarding few type of mutations. There is need to develop an algorithm which can annotate the mutations present in the complete genome and provide the complete information of the type of mutation and its site along with the possible drug compounds for treating the same.