6 Summary

In this thesis, the prime focus was to develop the platform and tools which can be helpful to the scientific community working in the field of peptide-based therapeutics. First, a unified platform in the form of meta-database was developed named "SATPdb", which hosts a total of 19192 peptides having diverse functions which were experimentally validated in different studies. In SATPdb, all the peptides were categorized in different functional categories using manual curation. The major functional categories represented by SATPdb are 1) antimicrobial, 2) antibacterial, 3) antiviral, 4) antifungal, 5) antihypertensive, 6) anticancer, 7) drug delivery vehicle, 8) antiparasitic, 9) toxic and 10) cell-cell communication. Peptides which were not covered in the above major functional categories were represented using a separate 'Miscellaneous' category. For each major functional category, appropriate sub-categories were also created. For example, anticancer peptides were further categorized into sub-functional categories 'antitumor' and 'antiangiogenic' wherein the peptides from the former sub-category tend to kill the tumor cells by their cytotoxic property while the latter sub-category tend to block the angiogenesis process (process of formation of new blood vessels) in developing tumor to block its source of nutrients. SATPdb also provides structural annotation of its peptides, which is a unique feature and was either missing or under-represented in most of the peptide databases. State-of-art methods were used to perform the structural annotation of 17,284 peptides and all the structures are also available for downloading from the web interface of SATPdb. Briefly, SATPdb can have following applications: i) searching ~22 peptide databases corresponding to 19192 peptides in one go; ii) Annotating functional

information of a peptide based on its matching with a peptide already present in SATPdb; iii) Finding whether the query peptide is novel or has a similar peptide in SATPdb; iv) Finding a peptide having more than one function (moonlighting peptide) or peptide having desired multiple functions; v) Peptide structures present in SATPdb can be further used in docking and MD studies from which structure-function relationship can be explored and new analogs can be designed.

The next objective of this thesis was to fill the gap between traditional peptidology and computational peptidology with respect to handling non-natural as well as modified peptides which play an important part in improving the therapeutic property of peptides. In this regards, we developed a computational tool "PEPstrMOD" to predict the structure of peptides having natural, non-natural as well as modified peptides. Although a very limited attempt in this regard was previously made by Beaufays et al., the method was available neither in the form of web server nor standalone. Moreover, the method was also limited in the sense that only ~19 different types of modified residues were presented in their method. In PEPstrMOD, we integrated 147 different types of modified residues from the FFNCAA library and 210 types of modified residues from SwissSideChain library as well as 32 frequently occurring PTMs from FFPTM library. Therefore, PEPstrMOD can predict the structure of peptides having diverse types of natural, non-natural or modified residues. We also benchmarked the PEPstrMOD algorithm on a dataset of 34 cyclic peptides as well as 501 modified peptides which had at least one modified residue. On cyclic peptides the performance of PEPstrMOD was comparable to PEP-FOLD algorithm. Generally, the performance of PEPstrMOD was very good on the short peptides. Even on the large peptides, the performance of PEPstrMOD was

reasonable. Briefly, using PEPstrMOD, following types of modifications can be incorporated in the peptide to annotate its structure: a) terminal modifications (Nterminal/Cterminal); b) stereo-chemical modifications (D-amino acids); c) peptide cyclization (N-to-C or disulfide bridge between cysteine residues); d) non-natural residues and e) PTMs.

Next, we focused on performing the structural analysis of ligand binding sites on peptides. Understanding the peptide-ligand interactions is important to develop drug candidates which can treat neurological disease like Alzheimer's disease. These interactions also find application in the development of drug-delivery systems or in diagnostics. Based on the composition analysis, it was observed that the residues Cys, Leu and Ile are highly preferred in the interactions with the ligand molecules. Physico-chemical properties revealed that interacting residues are generally hydrophobic and non-polar in nature. We also performed the benchmarking of different docking methods on peptide-ligand interactions. Docking software are widely used to study protein-ligand interactions. However, our focus was to study peptide-ligand interactions and we had a question in mind that whether the current state-of-art protein-ligand docking methods can accurately model the peptide-ligand interactions or do we need novel docking methods specifically for studying peptide-ligand interactions? Generally, peptides are more flexible than proteins and therefore lack a well-defined active site of interaction and therefore we anticipated that modeling peptide-ligand interactions using currently available docking methods might be a difficult problem. Therefore, we benchmarked the seven docking methods (which are used to model protein-ligand interactions) on a dataset of 57 peptide-ligand complexes extracted and filtered from PDB. We used seven widely

used docking methods which were (i) AutoDock, (ii) AutoDock Vina, (iii) DOCK 6, (iv) PLANTS, (v) rDock, (vi) GEMDOCK and (vii) GOLD. These docking methods are freely available to use for academic purpose except GOLD, which is commercial. In our benchmarking study, we observed that TOP pose of all the docking methods performed poor in successfully reproducing the crystallographic pose. Among the seven methods, AutoDock performed the best with 21.05 % success while all the other methods had very poor success rate. On the basis of BEST docking pose out of top 30 poses, PLANTS had highest success rate (43.86 %) followed by AutoDock Vina (40.35 %) and AutoDock (35.09 %). However, in terms of RMSD between crystallographic pose and BEST docking pose, AutoDock performed the best with an average RMSD of 3.816 Å. GOLD which is a very popular and successful method to model protein-ligand interactions, completely failed to model the peptide-ligand interactions with the success rate of 3.51 % and an average RMSD of 11.669 Å on 57 peptide-ligand complexes. Therefore, this benchmark paves the way for the development of novel docking algorithms and methods that can specifically model peptide-ligand interactions. It also provided important information that currently AutoDock method can be used to study these interactions until novel algorithms are developed specifically for studying peptide-ligand interactions.

Despite the considerable progress in the field of computational peptidology, much work needs to be done in the future. For instance, accurate structure prediction of chemically modified peptides remains a major challenge in front of the computational scientist. Although, an attempt has been made in the present study, the structure prediction of peptides having complex modifications is still not possible because of the unavailability of force fields. Therefore, the development of accurate structure prediction methods for

chemically modified peptides is the need of the hour. Another area where efforts need to be done is the development of tools for studying the interaction of peptides and small molecule ligands. Small peptides like CPPs are very popular for drug delivery by traversing biological membrane without causing significant membrane damage. CPPs can be used to deliver a variety of molecules ranging from small molecules to macromolecules like proteins and nucleic acids. Therefore, computational methods like docking, for predicting the interaction of CPPs with these cargoes will certainly be very helpful for designing appropriate formulations for drug delivery before carrying out wet experiments.