Chapter 5

Conclusions

In this thesis we tested GEMDOCK on molecular recognition and evaluated the application of virtual screening. GEMDOCK achieved a success rate of 78% on benchmark testing of the CCDC/Astex test set and we utilized this test set to make characteristics analysis of GEMDOCK. A virtual screening application of human α-thrombin was carried out on GEMDOCK and GEMDOCK achieved the enrichment of 20.2 with pharmacological preferences on this target. In addition we developed a novel strategy for integrating and improve the virtual screening accuracy. This method was to combine different docking method via fusion their ranking results and the analysis of this thesis proved it indeed useful in virtual screening. This fusion method could provide better accuracy and more efficiency than the widely used strategy, consensus scoring.

5.1 Summary

The performance of the robust, flexible docking method GEMDOCK in molecular docking and virtual screening is validated in this research. We analyzed the influence factors of GEMDOCK and found that performance of GEMDOCK was affected by size of receptor and number of single bonds. These factors would be increased the searching space of ligand conformation and reduce the accuracy of docking solutions. The successful rate of GEMDOCK on CCDC/Astex test set is 78% and improves to 83% when we retain structure water molecules such shown in Table 5. In general GEMDOCK and its scoring function can find correct solutions on different kinds of ligands for protein such as small and polar, large and polar, highly flexible, rigid and hydrophobic. But GEMDOCK is also within certain limits. When the crystal structure has defect or some atoms of the ligand have special interaction type which the force field of GEMDOCK does not consider, GEMDOCK maybe fail in finding correct solutions in these special conditions. GEMDOCK is also verified on virtual screening of human a-thrombin and the result shows this docking program has the ability to be a useful virtual screening tool. We identify several important pharmacological preferences and interactions on thrombin. If using these preferences

knowledge into our scoring function, the enrichment factor could achieve 20.2 and the false positive rate is 3.4% at true positive rate 100%. These results show GEMDOCK is useful on virtual screening and our result is 20 fold superior to GOLD on human a-thrombin.

Another focus of this thesis is the application of data fusion in virtual drug screening. We tested and analyzed this novel concept for scoring method fusion which combines different scoring functions via ranking combination among three virtual screening sets. According to our test set, we fused the result of various scoring functions by their ranking orders and we found the combination of a pair of scoring methods with lowest false positive rates obtained the maximization of enrichment and the minimization of false positive rate. Consensus scoring is a widely used strategy in this area and we also make the comparison between ranking fusion and consensus scoring. Performances of ranking combination and scoring combination will promote with combination of different scoring methods. In our test data, the best performance of ranking combination occurs when fusing two methods but the best performance of scoring combination is inconsistency among the number of combinations. Among three fusion experiments the best accuracies of ranking combination are indeed superior to scoring combination. Our analyses indicate the ranking fusion performs better than consensus scoring in most cases and the ranking combination of GEMDOCK-both and GOLD-new can always obtain most improvement in our test.

5.2 Major Contributions and Future Works

In short, the major contributions of this thesis can be summarized as following:

- Validation GEMDOCK on a large and diversity test set.
- Characteristics analysis of GEMDOCK in molecule recognition
- Application of GEMDOCK to virtual screening of human α-thrombin
- Development of a novel strategy via data fusion for improving the accuracy of virtual screening

There are several possible directions which we will pursue in future:

- The application of GEMDOCK on virtual screening for different protein targets
- Applying thrombin research to protease drug screening, such as factor Xa and SARS 3CL

protease

- Further investigating and more applications for data fusion in virtual screening post-analysis