

CONTENTS

Abstract (in Chinese)	I
Abstract	II
Acknowledgements	III
Contents	IV
List of Tables	VI
List of Figures	VII
Chapter 1. Introduction	01
1.1 Motivations and Purposes	01
1.2 Related Works	02
1.3 Thesis Overview	04
Chapter 2. Materials and Methods	06
2.1 Scoring Function	06
2.2 Mining Pharmacological Consensuses	08
2.3 Flexible Docking Search Method	10
2.3.1 Recombination Operators	11
2.3.2 Mutation Operators	12
Chapter 3. Validation and Application of GEMDOCK	14
3.1 Molecular Recognition	15
3.1.1 Test Data Set	15
3.1.2 The Proteins/Ligand Preparation and GEMDOCK Parameters	16
3.1.3 Overall Accuracy on 305 Complexes	17
3.1.4 Characteristics Analysis	18
3.2 Virtual Screening of Human α -thrombin	20
3.2.1 Introduction to Virtual Screening	21
3.2.2 Introduction to Human α -thrombin	22
3.2.3 Parameters of GEMDOCK	22
3.2.4 Target and Data Set Preparations	23
3.2.5 Screening Accuracy Analysis	25
Chapter 4. Application of Data Fusion in virtual screening	28
4.1 Introduction to Data Fusion	28

4.2 The Data Set Preparation	29
4.3 Methods of Data Fusion	31
4.4 Screening Accuracy Analysis	32
Chapter 5. Conclusions	36
5.1 Summary	36
5.2 Major contributions and future works	37
Tables	39
Figures	57
References	71
Appendix	1
A-1 Source Code of Contact	1



LIST OF TABLES

Table 1. Atom Formal Charge of GEMDOCK	39
Table 2. Atom Types of GEMDCOK	40
Table 3. The 305 Test Complexes	41
Table 4. Parameters of GEMDOCK for Molecular Docking Parameters	42
Table 5. GEMDOCK Validation Results for Different Lists	43
Table 6. Comparing GEMDOCK with Gold on Different Lists	44
Table 7. Parameters of GEMDOCK for Virtual Screening	45
Table 8. Ligand preferences evolved from known active ligands are used to screen lead compounds for the human α -thrombin inhibitor complexes	46
Table 9. Comparing GEMDOCK with GOLD with Respect to Docking Ten Ligands back into Reference Protein (1dwd)	47
Table 10. Comparison of GEMDOCK with GOLD on screening 1010 compounds with false positive rates (%)	48
Table 11. GEMDOCK and GOLD Screening Accuracies Using Different Combinations of Scoring Functions	49
Table 12. The Ranks of Ten Known Thrombin Inhibitors Using GEMDOCK and GOLD with Different Combinations of Scoring Functions	50
Table 13. Statistics of Overall Screening Accuracy for Fusion Screening Ranking of HSV-1 Thymidine Kinase (TK)	51
Table 14. Statistics of Overall Screening Accuracy for Fusion Screening Ranking of Human Dihydrofolate Reductase (DHFR)	52
Table 15. Statistics of Overall Screening Accuracy for Fusion Screening Ranking of Estrogen Receptor (ER)	53
Table 16. Statistics of Overall Screening Accuracy for Fusion Screening Scoring of HSV-1 Thymidine Kinase (TK)	54
Table 17. Statistics of Overall Screening Accuracy for Fusion Screening Scoring of Human Dihydrofolate Reductase (DHFR)	55
Table 18. Statistics of Overall Screening Accuracy for Fusion Screening Scoring of Estrogen Receptor (ER)	56

LIST OF FIGURES

Figure 1. The main steps of GEMDOCK for virtual database screening GEMDOCK results for four typical acceptable complexes	57
Figure 2. The linear energy function of pair-wise atoms for steric interactions (light line), hydrogen bonds (bold line), and electrostatic potential in GEMDOCK	58
Figure 3. GEMDOCK results for four typical acceptable complexes (i.e., the RMSD value < 2.0 ?)	59
Figure 4. The successful percentages of GEMDOCK for retaining (red bar) and removing (black bar) structure water molecules in CCDC/Astex test set	60
Figure 5. Comparison of successful rates between GEMDCOK and GOLD in different rotational bond levels	61
Figure 6. GEMDCOK results for four factors for unacceptable examples (i.e., RMSD value > 2.0 ?)	62
Figure 7. Affected factors analysis of GEMDOCK performance	63
Figure 8. Ten known human α -thrombin ligands	65
Figure 9. The binding-site pharmacological consensuses	66
Figure 10. GEMDOCK screening accuracies of thrombin	67
Figure 11. Average enrichment factor and average false positive rate of three virtual screening targets	68
Figure 12. Ranks and score curves for three virtual screening targets	69
Figure 13. Average enrichment factor and false positive rate comparison for ranking and scoring combinations	70