

Copyright  
by  
Robert Dean Malmstrom  
2011

**The Dissertation Committee for Robert Dean Malmstrom Certifies that this is the approved version of the following dissertation:**

**Discovering Dengue Drugs-Together: Grid Scale Virtual Screening with Mean Field Free Energy of Binding Rescoring**

**Committee:**

---

Werner Braun Ph.D., Chair

---

Stanly J. Watowich Ph.D., Supervisor

---

James C. Lee Ph.D.

---

Xiaodong Cheng Ph.D.

---

James M. Briggs Ph.D.

---

Dean, Graduate School

**Discovering Dengue Drugs-Together: Grid Scale Virtual Screening with  
Mean-Field Free Energy of Binding Rescoring**

**BY**

**Robert Dean Malmstrom, B.S.**

**Dissertation**

Presented to the Faculty of the Graduate School of

The University of Texas Medical Branch

in Partial Fulfillment

of the Requirements

for the Degree of

**Doctor of Philosophy**

**The University of Texas Medical Branch**

**August 2011**

## **DEDICATION**

For My Family

All of Them

## ACKNOWLEDGEMENTS

First and foremost I need to thank my mentor

*Stan Watowich*

And my committee

*Werner Braun, Jim Lee, Xiaodong Cheng, Jim Briggs*

The members of World Community Grid for volunteering their computers

The amazing staff of World Community Grid

Particularly

*Bill Bovermann, Keith Uplinger, Viktors Berstis, Al Seippel, Jonathan Armstrong*

The Texas Advanced Computing Center

Particularly

*Jay Boisseau, Dean Nobles*

Our collaborators and the developers of Mean Field Free Energy of Binding

*Benoit Roux, Yuqing Deng*

All current and past members of the Watowich Lab

Particularly

*Andrew Russo, Suzanne Tomlinson*

Others at UTMB

*Deb Botting, Lillian Chan, Sarita Sastry, Jim Halpert, Mary Moslen-Treinen, Mark White*

The Keck Center Pharmacoinformatics Training Program of the Gulf Coast Consortia  
(NIH Grant No. T90 DK071504-03)

My Family and Friends

*Elise, Bjorn, Daniel*

*Dean & Anick Malmstrom, Madeline Koman*

*Bill Rose, Alex Esadze, The Cornelius Family, Jon Read*

And My Heavenly Father.

# **Discovering Dengue Drugs-Together: Grid Scale Virtual Screening with Mean-Field Free Energy of Binding Rescoring**

Publication No. \_\_\_\_\_

Robert Dean Malmstrom, Ph.D.

The University of Texas Medical Branch, 2011

Supervisor: Stanly Watowich

The virtual screening of chemical databases against drug discovery targets with docking programs can enrich a database for bioactive compounds. However, current virtual screening methods generate many false positives leading to extensive and expensive testing of ultimately inactive compounds. In addition, the performance of virtual screening methods is dependent on the target system. This work examines if coupling traditional docking based virtual screening methods with perturbation based mean field free energy of binding (MF-FEB) calculations to rescore docking generated poses will improve enrichment over traditional virtual screening methods. MF-FEB calculations are computationally demanding requiring the distributed computing resources of IBM's World Community Grid. The work details three retrospective studies of MF-FEB using a 30 compound test set for binders and non-binders of the L99A T4 lysozyme, the DUD estrogen agonist test set and the DUD trypsin test set. In addition, this work describes the active prospective drug discovery effort on World Community Grid Discovering Dengue Drugs-Together that utilizes MF-FEB rescoring. The testing MF-FEB rescoring showed that while MF-FEB calculations can improve enrichment over traditional virtual screening methods it still has many of the same limitations as traditional virtual screening methods.

## TABLE OF CONTENTS

|  |             |
|--|-------------|
| <b>LIST OF TABLES</b>  | <b>XIII</b> |
| <b>LIST OF FIGURES</b>   | <b>XV</b>   |
| <b>LIST OF COMMON ABBREVIATIONS</b>                                  | <b>XVII</b> |
| <b>CHAPTER 1 CHALLENGES IN COMPUTER-AIDED DRUG DISCOVERY</b>         | <b>1</b>    |
| 1.1 Computational Methods in Drug Discovery .....                    | 2           |
| 1.1.1 Brief Overview of Drug Discovery .....                         | 2           |
| 1.1.2 Computers in Drug Discovery .....                              | 5           |
| 1.1.3 Predicting Binding .....                                       | 7           |
| 1.2 Docking Programs .....   | 9           |
| 1.2.1 Scoring Functions .....  | 9           |
| 1.2.1.1 Force Field Based Scoring .....                              | 10          |
| 1.2.1.2 Empirical Based Scoring Functions .....                      | 14          |
| 1.2.1.3 Consciences Scoring Methods .....                            | 17          |
| 1.2.2 Search Methods .....   | 17          |
| 1.2.2.1 Iterative Search Methods .....                               | 18          |
| 1.2.2.2 Stochastic Search Methods .....                              | 20          |
| 1.2.2.3 Simulation Based Search Methods .....                        | 21          |
| 1.2.2.4 Combining Search Methods .....                               | 22          |
| 1.2.3 Docking Performance .....                                      | 23          |
| 1.2.4 Example Program: AutoDock4 .....                               | 25          |
| 1.3 Virtual Screening .....  | 28          |
| 1.3.1 Virtual Chemical Libraries .....                               | 28          |
| 1.3.2 Virtual Screening Performance .....                            | 30          |
| 1.4 Statistical Mechanics Based Free Energy of Binding Methods ..... | 32          |
| 1.4.1 End Point Methods .....  | 33          |
| 1.4.1.1 LIE .....  | 33          |
| 1.4.1.2 MM-PBSA .....  | 34          |
| 1.4.2 Pathway Methods – Free Energy Perturbation .....               | 36          |

|  |  |           |
|--|--|-----------|
| 1.4.3  | Application in Drug Discovery .....                | 38        |
| 1.4.4  | Mean Field Free Energy of Binding .....            | 38        |
| 1.5  | Introduction to High Performance Computing .....   | 42        |
| 1.5.1  | Types of High Performance Computers .....          | 42        |
| 1.5.2  | Drug Discovery with Grid Computing .....           | 43        |
| 1.6  | Overview of Work .....                             | 44        |
| <b>CHAPTER 2 METHODS FOR DOCKING, VIRTUAL SCREENING, AND MEAN<br/>FIELD-FREE ENERGY OF BINDING</b> |  | <b>47</b> |
| 2.1  | Computational Resources .....                      | 47        |
| 2.1.1  | Software .....                                     | 47        |
| 2.1.1.1  | Docking .....                                      | 47        |
| 2.1.1.2  | Molecular Simulation .....                         | 48        |
| 2.1.1.3  | Molecular Visualization and Manipulation .....     | 48        |
| 2.1.1.4  | Statistics .....                                   | 49        |
| 2.1.1.5  | Scripting .....                                    | 49        |
| 2.1.2  | Hardware .....                                     | 49        |
| 2.1.3  | Molecular File Types .....                         | 49        |
| 2.1.3.1  | The .pdb File Format .....                         | 50        |
| 2.1.3.2  | The .mol2 File Format .....                        | 51        |
| 2.2  | Preparing the Ligand and the Receptor .....        | 51        |
| 2.2.1  | Ligand Prep .....                                  | 52        |
| 2.2.1.1  | Obtaining Ligand Coordinates .....                 | 52        |
| 2.2.1.2  | Preparing Ligand .mol2 .....                       | 54        |
| 2.2.2  | Target Prep .....                                  | 55        |
| 2.2.2.1  | Preparing the .pdb File .....                      | 55        |
| 2.2.2.2  | Correcting Missing Atoms .....                     | 56        |
| 2.2.2.3  | Chain Termini .....                                | 56        |
| 2.2.2.4  | Protonation .....                                  | 57        |
| 2.3  | Docking and Virtual Screening with AutoDock4 ..... | 57        |
| 2.3.1  | Docking with AutoDock4 .....                       | 57        |
| 2.3.1.1  | Stage 1: Ligand and Receptor File Generation ..... | 57        |

|   |   |           |
|---|---|-----------|
| 2.3.1.2   | Stage 2: Scoring Grid Generation .....  | 59        |
| 2.3.1.3   | Stage 3: Docking Simulation .....   | 60        |
| 2.3.1.4   | Stage 4: Results Analysis and Visualization.....  | 61        |
| 2.3.2   | Virtual Screening with AutoDock4 .....  | 61        |
| 2.3.2.1   | Compound Databases – Source and Generation.....   | 61        |
| 2.3.2.2   | Virtual Screening at TACC.....  | 62        |
| 2.3.2.3   | Results Processing .....  | 63        |
| 2.4   | MF-FEB Calculations .....   | 64        |
| 2.4.1   | Target and Ligand Preparation.....  | 64        |
| 2.4.1.1   | Ligand Preparation.....   | 64        |
| 2.4.1.2   | Target Preparation.....   | 65        |
| 2.4.1.3   | Solvent .....   | 66        |
| 2.4.2   | MF-FEB Calculations .....   | 66        |
| 2.4.2.1   | Site .....  | 67        |
| 2.4.2.2   | Bulk.....   | 73        |
| 2.4.2.3   | Results.....  | 75        |
| 2.4.3   | Scripts for MF-FEB Calculations on TACC.....  | 76        |
| 2.4.3.1   | Building the Jobs.....  | 76        |
| 2.4.3.2   | Running and Processing the Results Collection .....                                       | 78        |
| <b>CHAPTER 3     OPTIMIZING VIRTUAL SCREENING AND MEAN FIELD – FREE</b> |   |           |
| <b>ENERGY OF BINDING CALCULATIONS</b>                                   |   | <b>80</b> |
| 3.1   | Virtual Screening .....   | 80        |
| 3.1.1   | Docking Parameters for Virtual Screening.....   | 80        |
| 3.1.1.1   | Trypsin Test Set .....  | 81        |
| 3.1.1.2   | Pose Selection .....  | 83        |
| 3.1.1.3   | AutoDock4 Search Parameters for Virtual Screening .....                                   | 89        |
| 3.1.2   | Filtering Virtual Screening Results for MF-FEB Rescoring on World<br>Community Grid ..... | 93        |
| 3.1.2.1   | Methods and Test Set.....   | 93        |
| 3.1.2.2   | Results and Discussion .....  | 95        |
| 3.2   | MF-FEB Calculations .....   | 100       |
| 3.2.1   | Local Run MF-FEB Optimization .....   | 100       |

|  |  |            |
|--|--|------------|
| 3.2.1.1  | Methods.....   | 100        |
| 3.2.1.2  | System Size.....   | 101        |
| 3.2.1.3  | Perturbation Sampling Time.....  | 102        |
| 3.2.1.4  | Variation in Perturbation Stage Free Energies.....                                   | 103        |
| 3.2.2  | Optimization of GCMC.....  | 106        |
| 3.2.2.1  | Methods.....   | 106        |
| 3.2.2.2  | GCMC $\mu$ Scan.....   | 106        |
| 3.2.2.3  | GCMC Cycles.....   | 108        |
| <b>CHAPTER 4 USING FREE ENERGY OF BINDING CALCULATIONS TO IMPROVE THE ACCURACY OF VIRTUAL SCREENING PREDICTION</b> |  | <b>112</b> |
| 4.1  | General Methods.....   | 112        |
| 4.1.1  | Hardware.....  | 113        |
| 4.1.2  | Atomic Coordinates.....  | 113        |
| 4.1.3  | Docking Calculations and MF-FEB Calculations.....                                    | 113        |
| 4.2  | Accuracy of MF-FEB Calculations Using AutoDock4-Predicted Binding Poses.....         | 114        |
| 4.2.1  | Test Set and MF-FEB Calculations.....  | 114        |
| 4.2.2  | Results and Discussion.....  | 115        |
| 4.3  | Enrichment and Discrimination Characteristics of MF-FEB Based Virtual Screening..... | 118        |
| 4.3.1  | Test Set and MF-FEB Calculations.....  | 119        |
| 4.3.2  | Results and Discussion.....  | 119        |
| 4.4  | Conclusions.....   | 125        |
| <b>CHAPTER 5 DISCOVERING DENGUE DRUGS-TOGETHER AND INFLUENZA ANTIVIRAL DRUG SEARCH</b>                             |  | <b>127</b> |
| 5.1  | Virtual Screening with AutoDock4 and MF-FEB Rescoring.....                           | 127        |
| 5.1.1  | Phase 1.....   | 128        |
| 5.1.2  | Phase 2.....   | 128        |
| 5.2  | World Community Grid.....  | 129        |
| 5.3  | Discovering Dengue Drugs-Together.....   | 131        |
| 5.3.1  | Project Goals.....   | 131        |

|  |            |
|--|------------|
| 5.3.2 Virtual Screening Targets .....  | 132        |
| 5.3.3 Current Progress Report.....   | 133        |
| 5.4 Influenza Antiviral Drug Search.....   | 133        |
| 5.4.1 Project Goals.....   | 133        |
| 5.4.2 Current Progress Report.....   | 133        |
| <b>CHAPTER 6 GRID IMPLEMENTATION OF VIRTUAL SCREENING AND MEAN<br/>FIELD-FREE ENERGY OF BINDING CALCULATIONS</b> | <b>134</b> |
| 6.1 Phase 1 – Virtual Screening.....   | 134        |
| 6.1.1 Preparing for Phase 1 Virtual Screening.....   | 134        |
| 6.1.1.1 Building the Compound Database .....   | 134        |
| 6.1.1.2 Target Preparation.....  | 135        |
| 6.1.1.3 AutoGrid4 and AutoDock4 Parameters .....   | 136        |
| 6.1.2 Virtual Screening on World Community Grid.....   | 136        |
| 6.1.2.1 Preparing Work for and Submitting Work to World<br>Community Grid .....                                  | 136        |
| 6.1.2.2 Docking on World Community Grid .....  | 137        |
| 6.1.2.3 Processing Results .....   | 139        |
| 6.1.3 Virtual Screening Run Times on World Community Grid.....   | 140        |
| 6.1.4 Results Extraction .....   | 140        |
| 6.2 Phase 2 – Grid Based MF-FEB Calculations.....  | 141        |
| 6.2.1 Work Preparation .....   | 141        |
| 6.2.1.1 Preparing Ligands .....  | 141        |
| 6.2.1.2 Preparing Target.....  | 142        |
| 6.2.1.3 MF-FEB CHARMM Scripts and Parameter Files .....  | 142        |
| 6.2.2 Running MF-FEB on World Community Grid.....  | 143        |
| 6.2.2.1 Modifying CHARMM Inputs .....  | 143        |
| 6.2.2.2 Performing MF-FEB Calculation .....  | 144        |
| 6.2.2.3 MF-FEB Job Creation and Management Scripts .....   | 147        |
| 6.2.3 MF-FEB Grid Performance .....  | 152        |
| 6.2.3.1 Runtime.....   | 152        |
| 6.2.3.2 Job Failure Rates.....   | 152        |

|                  |  |            |
|------------------|--|------------|
| <b>CHAPTER 7</b> | <b>EFFICIENCY AND EFFICACY OF RESCORING DOCKING GENERATED POSES BY PERTURBATION FREE ENERGY OF BINDING CALCULATION IN LARGE DATASETS</b> | <b>156</b> |
| 7.1              | Methods.....   | 157        |
| 7.1.1            | DUD Test Sets .....  | 157        |
| 7.1.2            | Virtual Screening .....  | 158        |
| 7.1.3            | MF-FEB Rescoring.....  | 158        |
| 7.2              | Enrichment.....  | 158        |
| 7.3              | Sources of Error in the MF-FEB Calculations.....   | 167        |
| 7.3.1            | Pose and Ranking.....  | 167        |
| 7.3.2            | The Relationships between MF-FEB Energetic Components .....  | 169        |
| <b>CHAPTER 8</b> | <b>CONCLUSIONS AND FUTURE DIRECTIONS</b>   | <b>176</b> |
| 8.1              | MF-FEB Rescoring as a Tool for Drug Discovery .....  | 176        |
| 8.2              | Improving Discovering Dengue Drugs-Together .....  | 178        |
| 8.2.1            | Improvements to the Past.....  | 178        |
| 8.2.2            | Best Practices for Virtual Screening with MF-FEB Rescoring .....   | 180        |
| 8.3              | Future Directions .....  | 182        |
| 8.3.1            | Test systems .....   | 184        |
| 8.3.2            | Methods.....   | 186        |
| 8.3.3            | Potential Results and Conclusions .....  | 188        |
| 8.4              | Improving Drug Discovery Tools.....  | 190        |
|                  | <b>BIBLIOGRAPHY</b>  | <b>192</b> |
|                  | <b>VITA</b>  | <b>202</b> |

## List of Tables

|   |     |
|---|-----|
| Table 3-1 – The Trypsin Test Set (*Number of bonds allowed to rotate in AutoDock4 docking simulations).....   | 82  |
| Table 3-2 – The RMSD (Å) of AutoDock4 generated poses, selected using a variety of criteria, to the experimentally determined poses. ....   | 87  |
| Table 3-3 – Changes in RMSD between docking generated and experimentally determined ligand pose at different maximum number energy evaluations parameter settings. ....   | 88  |
| Table 3-4 – Average RMSD of AutoDock4 generated pose vs. experimental determined poses and average runtime for the trypsin test set using different docking parameters. ....  | 91  |
| Table 3-5 – Comparison of “Gold” and “Production” docking parameters average RMSD of AutoDock4 generated pose vs. experimental determined poses and average runtime for each compound in the trypsin test set. n=3..... | 92  |
| Table 3-6 – Selected DUD targets. ....  | 94  |
| Table 3-7 – Population statistics for score and clusters size for each DUD virtual screening. ....  | 99  |
| Table 3-8 – Comparison of free energy of binding calculations using MF-FEB for benzene binding to L99A T4 lysozyme with different solvent drop radii. n=4..   | 101 |
| Table 3-9 – Comparison of free energy of binding calculations using MF-FEB for benzene binding to L99A T4 lysozyme with two different solvent drop sizes and two different perturbation MD run times. n=4 .....         | 103 |
| Table 3-10 – Comparison of four MF-FEB calculation on benzene binding to L99A T4 lysozyme. ....   | 105 |
| Table 4-1 – Comparison between MF-FEB calculations and experimental binding energies using inhibitor-lysozyme co-crystal structures as the starting poses for   |     |

the  $\Delta\Delta G_{\text{binding}}$  calculations. MF-FEB calculations were performed in triplicate.

\*Root-mean-square deviation (RMSD) reflects the displacement of heavy atom positions between final equilibrated pose and the starting structure. .... 115

Table 4-2 – MF-FEB predictions performed with self-docked poses. The root-mean-square deviation (RMSD) listed in the docking column compared AutoDock4-generated poses relative to the co-crystal structures. Only non-hydrogen atoms were used to tabulate RMSD values. The MF-FEB calculations were repeated 3 times with different starting seeds. The experimental binding data is repeated for ease of reference. .... 116

Table 4-3 – MF-FEB prediction using an apo lysozyme protein structure as the docking target. For the MF-FEB calculations, three replicated calculations were performed. The experimental binding data is repeated for ease of reference.... 118

Table 4-4 – Small-scale virtual screening results for all 30 ligand comparing experimental FEB to docking score and MF-FEB. NB= Non-binder (experimentally determined)..... 124

Table 7-1 –  $\Delta\Delta G_{\text{binding}}$  and rank order from MF-FEB calculations using different starting poses..... 168

Table 7-2 – Discrimination between binders and decoys for  $\Delta\Delta G_{\text{binding}}$  energy components for the ERAG test set. Highlighted cell have a p-score less than 0.05. .... 171

Table 7-3 – Discrimination between binders and decoys for  $\Delta\Delta G_{\text{binding}}$  energy components for the TRYP test set. Highlighted cell have a p-score less than 0.05. .... 172

Table 7-4 – Heat map of correlation between FEB energy components for the ERAG test set..... 174

Table 7-5 – Heat map of correlation between FEB energy components for the TRYP test set..... 175

## List of Figures

- Figure 2-1 – The “site” (left) and “bulk” (right) systems used for MF-FEB calculations for the binding of benzene to the L99A T4 lysozyme engineered hydrophobic binding site. The red spheres represent the locations of water molecules. Some waters were removed to show ligand and target. The mint green ribbon represents the backbone of the target. The stick model in the center represents benzene with green for carbon atoms and white for hydrogen atoms..... 68
- Figure 3-1 – Scatter plots of docking results by docking score and cluster size for five DUD test sets. Yellow squares are binders and blue diamonds are decoys..... 96
- Figure 3-2 – The change in the number for waters in the site system over 25 GCMC cycles at different  $\mu$  values ( $m = \mu$ ). Each GCMC checkpoint is 1/10 of a GCMC cycle. nMax corresponds to the maximum water that could be added. .... 110
- Figure 3-3 – Water count as a function of GCMC cycle for two systems with the first 10 cycles highlighted. .... 111
- Figure 4-1 – Enrichment curve showing docking-MFFEBS enrichment compared to docking only enrichment in a virtual screening on a 30 member compound base with 16 known binders and 14 known non-binders. The ideal line represent the best case result where no false positives. The random line corresponds with a random distribution of binders and non-binders. .... 120
- Figure 4-2 – Energy distribution of ligands rank ordered by docking score (circles) or MF-FEB energy (squares). Filled and unfilled markers represent lysozyme binders and non-binders, respectively..... 122
- Figure 5-1 – Work flow diagram for virtual screening with AutoDock4 and MF-FEB rescoring..... 129
- Figure 6-1 – MF-FEB calculation cumulative success and failure rates by stage on World Community Grid for ERAG test set..... 155

|  |     |
|--|-----|
| Figure 6-2 – MF-FEB calculation cumulative success and failure rates by stage on World Community Grid for TRYP test set. ....  | 155 |
| Figure 7-1 – Plot of DUD’s estrogen receptor agonist test set enrichment curves sorted by AutoDock4’s scoring function and MF-FEB rescoring. Dashed lines correspond to ideal enrichment and random enrichment.....                                      | 162 |
| Figure 7-2 – Plot of DUD’s trypsin test set enrichment curves sorted by AutoDock4’s scoring function and MF-FEB rescoring. Dashed lines correspond to ideal enrichment and random enrichment.....  | 163 |
| Figure 7-3 – Success rates plot for DUD’s estrogen receptor agonist test set. All bars start at 0% and are overlaid for comparison. Ideal corresponds to the total number of binders in the test set. ....   | 164 |
| Figure 7-4 – Success rates plot for DUD’s trypsin test set. All bars start at 0% and are overlaid for comparison. Ideal corresponds to the total number of binders in the test set. There were no binders found using AutoDock4’s scoring function. .... | 164 |
| Figure 7-5 – Overlay of MF-FEB rescoring enrichment curve on enrichment curves from other docking programs for DUD’s estrogen receptor agonist test set. (Cross, Thompson et al. 2009) .....   | 165 |
| Figure 7-6 – Overlay of MF-FEB rescoring enrichment curve on enrichment curves from other docking programs for DUD’s trypsin test set. (Cross, Thompson et al. 2009) .....   | 166 |

## List of Common Abbreviations

|                   |  |
|-------------------|--|
| ADT .....         | AutoDock Tools                           |
| CADD .....        | Computer-Aided Drug Discovery            |
| CHARMM .....      | Chemistry at HARvard Molecular Mechanics |
| CPU .....         | Central Processing Unit                  |
| DDDT .....        | Discovering Dengue Drugs-Together        |
| DUD .....         | Database of Useful Decoys                |
| ERAG .....        | Estrogen Receptor Agonist DUD test set   |
| FEB .....         | Free Energy of Binding                   |
| FEP .....         | Free Energy Perturbation                 |
| GA .....          | Genetic Algorithm                        |
| GAFF .....        | Generalized Amber Force Field            |
| GCMC .....        | Grand Canonical Monte Carlo simulations  |
| GNU .....         | GNU's Not Unix                           |
| GSBP .....        | General Solvent Boundary Potential       |
| H-bond(ing) ..... | Hydrogen Bond(ing)                       |
| HCV .....         | Hepatitis C Virus                        |
| HIV .....         | Human Immunodeficiency Virus             |
| HIVP .....        | HIV-1 Protease DUD test set              |
| HIVR .....        | HIV-1 Reverse Transcriptase DUD test set |
| IADS .....        | Influenza Antiviral Drug Search          |
| LIE .....         | Linear Interaction Energy                |
| MC .....          | Monte Carlo simulations                  |
| MD .....          | Molecular Dynamics simulations           |
| MF-FEB .....      | Mean Field Free Energy of Binding        |

|         |  |
|---------|--|
| MM      | Molecular Mechanics                                    |
| MM-PBSA | Molecular Mechanics and Poisson-Boltzmann Surface Area |
| MNEE    | Maximum Number of Energy Evaluations                   |
| NADU    | NeurAminidase DUD test set                             |
| NS      | Non-Structural   |
| PBC     | Periodic Boundary Condition                            |
| PB-FEB  | Perturbation Based Free Energy of Binding              |
| PDB     | Protein Data Bank                                      |
| PFF     | Polarizable Force Fields                               |
| PMF     | Potential Mean Force                                   |
| QSAR    | Qualitative Structure Activity Relationship            |
| RAM     | Random Access Memory                                   |
| RMSD    | Root Mean Squared Deviation                            |
| SBDD    | Structure Based Drug Discovery/Design                  |
| SM      | Statistical Mechanics                                  |
| SSBP    | Solvent Shell Boundary Potential                       |
| TACC    | Texas Advanced Computing Center                        |
| TRYP    | TRYPSin DUD test set                                   |
| VdW     | Van der Waal's   |
| WHAM    | Weighed Histogram Analysis Method                      |
| ZINC    | ZINC Is Not Commercial                                 |

## Chapter 1 Challenges in Computer-Aided Drug Discovery

Drug discovery and development is an expensive and challenging process. Over the past ten years the cost of drug research and development has approximately doubled while the average number of Food and Drug Administration (FDA) approved drugs is largely unchanged. (Paul, Mytelka et al. ; Service 2004; Aronovitz 2006) The high cost of drug discovery has led to the so called “Block Buster” syndrome in which corporate drug discovery efforts focus on finding treatments for diseases with high profit values such as those that require continuing treatment or that have a high occurrence in the developed world. (Mrazek and Mossialos 2003; Service 2004; Stirner 2008) The effect of the “Block Buster” syndrome is that many serious diseases in the developing world are not targeted for drug development. (Mrazek and Mossialos 2003; Service 2004; Stirner 2008) As an alternative to traditional drug discovery methods, computational drug discovery tools provide a potentially lower cost approach to drug discovery. Lower drug discovery costs allow academic and charitable organization to perform drug discovery projects focused on diseases with a greater humanitarian burden than profitability, such as malaria, leishmania, and dengue fever. Additionally, lower drug discovery costs allow for increased corporate efficiency in drug development thereby increasing treatment options. (Trouiller, Olliaro et al. 2002; Nwaka and Ridley 2003; Stirner 2008)

A good computational drug discovery tool should identify bioactive compounds from a chemical library, in a variety of different systems, in a timely and efficient manner. The goal of this work is to improve computational drug discovery tools by improving the success rate in docking based virtual screening, focusing on the prediction of binding. Improved success rates reduces the total number of compounds tested at the bench top, thus decreasing drug discovery cost and increasing the probability of lead generation and future drug development.

This chapter introduces computer-aided drug discovery, or CADD, focusing on the computational approaches for predicting the binding of small drug-like molecules to protein

targets. Section 1.1 of this chapter summarizes the ways computational methods are integrated into drug discovery and development. Section 1.2 focuses on docking programs, the tools for binding prediction, and explains different search methods and scoring functions used to make binding predictions. Section 1.3 describes the effectiveness of virtually screening chemical libraries using docking programs. Section 1.4 introduces statistical mechanics based methods for the calculation of free energy of binding (FEB), and discusses its uses as an alternative to the docking programs' scoring functions. Section 1.5 is a brief introduction to high performance computing and its use in drug discovery. Finally, Section 1.6 states our hypothesis and outlines the rest of this work.

## **1.1 COMPUTATIONAL METHODS IN DRUG DISCOVERY**

This section provides a general overview of CADD. Section 1.1.1 outlines the process and challenges of how a chemical compound becomes a drug. Section 1.1.2 introduces CADD methods and how they are used in each stage of the drug discovery process. Finally, Section 1.1.3 concludes with an introduction to computational methods used to predict the free energy of binding (FEB).

### **1.1.1 Brief Overview of Drug Discovery**

Drugs are generally small organic molecules (< 500 Da) (Lipinski, Lombardo et al. 2001) which agonizes, antagonizes, or inhibits a target molecule in a biological pathway producing a physiological response. Most drugs targets are proteins, and the function of the target proteins vary from receptors, to enzymes, to ion channels. (Drews 2000)

The process of drug discovery and development is the identification, or creation, of a bioactive compound and its subsequent modification in to a drug, balancing both its potency and safety. (Barril, Brough et al. 2005; Mitscher and Dutta 2006) To discover and develop a new drug can take upwards of 10 to 15 years. (Lipsky and Sharp 2001; Aronovitz 2006) The path of drug discovery and development has four stages. (Baxter and Lockey 2001; Bleicher, Bohm et

al. 2003; Nwaka and Ridley 2003; Deprez-Poulain and Deprez 2004; Aronovitz 2006; Keseru and Makara 2006; Mitscher and Dutta 2006)

The first stage in drug discovery is target selection. (Baxter and Lockey 2001; Bleicher, Bohm et al. 2003; Nwaka and Ridley 2003; Deprez-Poulain and Deprez 2004; Aronovitz 2006; Keseru and Makara 2006; Mitscher and Dutta 2006) Target selection starts with the basic science research required to understand the pathology of the disease to determine possible pharmacological interventions. Once a target is selected, the data obtained on the target determines how drug discovery will proceed. If there is an experimental structure of the target, then structure based drug discovery (SBDD) is used, where the discovery process is governed by the shape and chemistry of the target. In other cases, if the structure of the target is unknown, but the structure of a ligand is known (Andricopulo, Salum et al. 2009), then ligand based drug discovery is used, where drug discovery proceeds is based on modifying the structure and chemistry of a known ligand.

The goal of the second stage of drug discovery is hit generation. Hits are compounds that are “active” against the target. (Baxter and Lockey 2001; Bleicher, Bohm et al. 2003) Hits are identified through methods such as high throughput screening (HTS) (Macarron 2006), fragment screening (Fischer and Hubbard 2009), or virtual screening (Shoichet 2004). In these methods, researchers screen chemical libraries to identify “active” compounds. The activity of a compound is based on some predetermined threshold of potency, like binding or inhibition concentrations. (Mitscher and Dutta 2006) A common metric for classifying a compound as a hit is that the compound has activity at low micromolar concentration. (Baxter and Lockey 2001)

In the third stage of drug discovery, researchers develop hits into lead compounds and then into candidate compounds for human trials. Lead compounds are chemically modified hits with increased potency, usually in the high (>100) nanomolar range. (Baxter and Lockey 2001; Bleicher, Bohm et al. 2003) The lead compound’s absorption, distribution, metabolism,

excretion, and toxicity (ADME-Tox) are tested at this stage and the lead is modified to improve safety. (Hodgson 2001) A lead compound becomes a candidate compound for human trial when it achieves safety and potency requirements. (Baxter and Lockey 2001; Bleicher, Bohm et al. 2003)

The final stage of drug discovery is the process of the candidate compound becoming a drug through clinical trials. (Lipsky and Sharp 2001; Aronovitz 2006) Clinical trials in the United States have three stages. The first stage determines the drug's safety in healthy adults. The second proves potency of the drug in a small study with hundreds of subjects. The third stage tests safety and potency of the drug in a large study with thousands of subjects. If a drug passes all three stages, it then goes to market, during which its potency and safety is continuously monitored. (Lipsky and Sharp 2001; Aronovitz 2006)

The probability of a hit becoming a candidate is significantly less than one percent, and of the candidate compounds that proceed to clinical trials four in five will fail. (Paul, Mytelka et al. ; Baxter and Lockey 2001; Bleicher, Bohm et al. 2003; Gershell and Atkins 2003; Deprez-Poulain and Deprez 2004; Aronovitz 2006; Federsel 2006) Reviews of the drug discovery process identify lead development as the weak link in the process. (Paul, Mytelka et al. ; Baxter and Lockey 2001; Hodgson 2001; Bleicher, Bohm et al. 2003; Gershell and Atkins 2003; Deprez-Poulain and Deprez 2004; Aronovitz 2006; Federsel 2006) An FDA report states the problem as follows:

“For medical technology, performance is measured in terms of product safety and effectiveness. Not enough applied scientific work has been done to create new tools to get fundamentally better answers about how the safety and effectiveness of new products can be demonstrated, in faster time frames, with more certainty, and at lower costs. In many cases, developers have no choice but to use the tools and concepts of the last century to assess this century's candidates. As a result, the vast majority of investigational products that enter clinical trials fail.” (Aronovitz 2006)

The failure of candidate compounds in clinical trials is very expensive leading to an overall decrease in the number of compounds entering clinical trials, even as budgets for research and

development have increased. (Paul, Mytelka et al. ; Service 2004; Aronovitz 2006) Many solutions have been suggested to make drug discovery and development more effective including improved testing methods (Bleicher, Bohm et al. 2003; Deprez-Poulain and Deprez 2004), parallel potency and toxicity testing (Baxter and Lockey 2001), integrate production chemistry requirements in compound synthesis methods during lead development (Federsel 2006), and improving computational tools (Bleicher, Bohm et al. 2003). Fundamentally, the challenge is to predict accurately the chemical and biological properties of lead compounds.

### **1.1.2 Computers in Drug Discovery**

Computational tools play a central role in the pre-clinical design and development of drugs. These tools come from a variety of scientific disciplines, such as chemistry, biology, physics and mathematics. The collective employment of these tools in drug discovery is known as computer aided drug discovery/design/development or simply CADD. The utilization of CADD tools has led to the identification of numerous bioactive compounds and new drugs, most famously saquinavir and related human immunodeficiency virus-1 (HIV-1) protease inhibitors. (Talele, Khedkar et al. 2010) Each stage of pre-clinical drug development utilizes a different set of CADD tools.

The first stage of drug discovery, target selection, uses the computational tools developed in the fields of bioinformatics and structural biology. Bioinformatics uses statistical based relationship determining methods to elucidate the complex genetic and molecular interactions in cells from data obtained through high-throughput *in vivo* and *in vitro* methods (i.e. gene array, etc.) and meta-studies of scientific publications. (Kitano 2002; You 2004) These complex relationships can identify potential drug targets and predict adverse interactions. (Bhogal and Balls 2008; Schrattenholz and Soskic 2008) Once targets are identified, computational methods play a central role in the determination and modeling of the target's structure using data obtained from x-ray crystallography or NMR spectroscopy. (Anderson 2003; Villoutreix, Eudes et al. 2009) Alternately, if a target's structure cannot be determined experimentally, it can be modeled

from the macromolecular sequence alone, or by using homology modeling if similar structures have been determined. (Takeda-Shitaka, Takaya et al. 2004; Grant 2009) While not ideal, modeled targets are proving effective in drug discovery. (Bissantz, Bernard et al. 2003)

The second stage of drug discovery, hit identification, makes use of computational methods to create, generate, or identify compounds that will interact with the selected target site. New compounds can be created rationally using the target structure to guide the synthesis of new compounds. (For example (Duff, Mudhivarthi et al. 2009)) This process is called de novo design and can be done computationally to generate a new compound by starting with a chemical scaffold and “growing” the compound by adding functional groups. (Jorgensen 2009) More commonly, the virtual screening of virtual chemical libraries identifies hits. (Jorgensen 2009; Villoutreix, Eudes et al. 2009) Virtual chemical libraries can contain the complete molecular structures (Zhou 2011), or can be composed of fragments that are later combined into a novel compound. (Guvench and MacKerell 2009) Virtual screening relies on docking programs to determine the pose, the geometric relationship between the ligand and the target, and to determine the score, the “fitness”, of that pose. (Campbell, Gold et al. 2003; Kitchen, Decornez et al. 2004; Mohan, Gibbs et al. 2005; Leach, Shoichet et al. 2006) In virtual screening, the docking scores are used to rank the compounds and identify hits. (Campbell, Gold et al. 2003; Kitchen, Decornez et al. 2004; Mohan, Gibbs et al. 2005; Leach, Shoichet et al. 2006)

The third stage of drug discovery and development, lead optimization, uses the tools of chemoinformatics to predict the activity and ADME-Tox properties of compounds. Researchers determine ADME-Tox properties by using a mathematical model describing the empirical relationship between a set of molecular descriptors and a given activity or property. Examples of molecular descriptors include: directly measured or calculated physical properties (e.g., molecular weight, formal charge, pKa, solubility); chemical properties of inter-molecular interaction (e.g., number H-bond donors/acceptors, hydrophobic/polar surface area); reactivity (e.g., electronegative, HOMO and LUMO locations and intensity); the structure of the molecule

(e.g., functional groups, atomic coordinates, patterns of atomic bonding); and data from biological assay (e.g., binding constants, toxicity). (Zhou ; Engel 2006) Statistical and pattern recognition methods like linear regression, principle component analysis, machine learning, and evolutionary algorithms determine the quantitative relationship between the descriptors and the properties. (Engel 2006; Michielan and Moro 2010; Zhou 2011) The process of using the relationships of chemical descriptors to predict an activity or a property is known as a quantitative structure activity/property relationship (QSAR or QSPR). (Engel 2006; Kortagere and Ekins 2010; Michielan and Moro 2010; Zhou 2011)

Central to all the CADD tools used for drug discovery and development is the ability to predict the pharmacological and biophysical properties of compounds. While CADD methods are successfully used (Talele, Khedkar et al. 2010), often their predictions are inaccurate, generating false positives (Stouch, Kenyon et al. 2003; Tetko, Bruneau et al. 2006; Warren, Andrews et al. 2006; Cross, Thompson et al. 2009; Kortagere and Ekins 2010; Zhang 2011). Accurate property predictions decrease the cost of drug development by guiding development efforts towards compounds most likely to become drugs, avoiding the costly development of compounds that will ultimately fail. (Jorgensen 2004; Jorgensen 2009; Zhang 2011) Improving the predictive abilities of CADD tools can therefore decrease the cost of drug discovery permitting effective drug discovery projects targeting neglected diseases.

### 1.1.3 Predicting Binding

The binding of a ligand to its target is the central property of SBDD. Reversible binding<sup>1</sup> is the non-covalent association of two chemical species; for this work, a small organic molecule, the ligand (L) and a protein target (P). The chemical equation for simple stoichiometric reversible binding in an aqueous solution is written as



---

<sup>1</sup> Binding in this work always refers to reversible binding.

where PL is the protein ligand complex and the *aq* subscript denotes that the species in aqueous solution. The binding equilibrium constant for this reaction is

$$K_b = \frac{[PL_{aq}]}{[P_{aq}][L_{aq}]} \quad 1-2$$

where  $K_b$  is the binding constant and  $[x]$  are the different equilibrium concentrations. The absolute free energy of binding is defined as

$$\Delta G_{binding}^{\circ} = -RT \ln(C^{\circ} K_b) \quad 1-3$$

where  $\Delta G$  is the change in Gibbs free energy (constant volume free energy),  $R$  is the ideal gas constant,  $T$  is the temperature of the system in Kelvin, and  $C^{\circ}$  is the standard concentrations (1M).

The computational prediction of binding is based on developing approaches to calculate the  $\Delta G_{binding}$ . QSAR methods can establish functions for  $\Delta G_{binding}$  as a weighted collection of chemical descriptors, but these functions lack robustness because they are limited to the chemical space defined by the training set. The FEB is the global minimum of an energetic landscape. In SBDD, the energetic landscape is defined as a function of the spatial relationship of the atoms in the ligand and target. There are a number of computational approaches to model the energy landscape and find the minimum. SBDD methods are potentially more robust than the empirical QSAR methods as long as the binding energy landscape can be accurately determined. However, as discussed below, while current SBDD methods are able to identify binding compounds from a database, they are inaccurate generating false positives. (Warren, Andrews et al. 2006; Cross, Thompson et al. 2009; Kortagere and Ekins 2010) This work therefore seeks to improve the computational methods for binding prediction in order to provide more effective drug discovery tools.

## 1.2 DOCKING PROGRAMS

Docking programs are computational SBDD tools used to predict the binding of a ligand to a target. (Campbell, Gold et al. 2003; Kitchen, Decornez et al. 2004; Mohan, Gibbs et al. 2005; Leach, Shoichet et al. 2006) A docking program is composed of two components: a scoring function, and a search method. The scoring function approximates the FEB landscape. The search method explores the energetic landscape defined by the scoring function, searching for the global minimum. A docking program therefore predicts the pose, the geometric relationship of the ligand to the target, and the score, the energetic relationship between the ligand and the target. (Halperin, Ma et al. 2002; Campbell, Gold et al. 2003; Kitchen, Decornez et al. 2004; Cummings, DesJarlais et al. 2005) Researchers optimize search methods and scoring functions of docking programs to explore the binding energy landscape as efficiently as possible. (Campbell, Gold et al. 2003) There are a variety of docking programs, each with different search methods and scoring functions. (Halperin, Ma et al. 2002; Campbell, Gold et al. 2003; Kitchen, Decornez et al. 2004; Mohan, Gibbs et al. 2005) Section 1.2.1 and 1.2.2 introduces docking by examining the different classes of scoring functions and search methods respectively. Section 1.2.3 evaluates the current performance of docking programs. Section 1.2.4 concludes with a detailed description of the AutoDock4 docking program.

### 1.2.1 Scoring Functions

Scoring functions describe the binding energy landscape that the search methods explore. The goal of scoring functions is to provide quantitative discernment among docking poses. There are two categories of scoring functions: force field based scoring functions, and empirical scoring functions. (Kitchen, Decornez et al. 2004; Mohan, Gibbs et al. 2005) Sections 1.2.1.1 describes force field based scoring functions and introduces force fields in general. Section 1.2.1.2 describes empirical scoring functions. Section 1.2.1.3 concludes with consensus scoring.

### ***1.2.1.1 Force Field Based Scoring***

Molecular mechanics force fields define the total potential energy of a system as a function of atomic position. (Kitchen, Decornez et al. 2004; Mohan, Gibbs et al. 2005; Guvench and MacKerell 2008) As the position of the atoms change relative to each other the potential energy of the system changes. Researchers first used force fields to simulate a variety of chemical phenomena starting with the properties of the ideal gases. Later, condensed phase simulations lead to the development of force fields used to stimulate proteins and other macromolecules. Force fields are used for docking as well as for free energy of binding calculations, as described below. This section is divided into two parts: the first describing force fields generally, and the second describing specific uses of force fields as scoring functions.

#### *1.2.1.1.1 Force Fields in General<sup>2</sup>*

The potential energy of a system can be determined as a function of the position of each body in that system in relation to the force field acting upon it. The way that the bodies and the forces are defined differentiates molecular mechanics force fields. The granularity of a force field refers to how it models each body, from atomistic fine-grain models, to course-grain models where each body may represent a chemical functional group or an amino acid. The phenomena being studied, as well as the computational power available, determines the granularity used for simulations. For docking, researchers use fine grain atomistic force fields originally designed to

---

<sup>2</sup> In this section the generic force field equations were taken from a review by Guvench, O. and A. D. MacKerell, Jr. (2008). "Comparison of protein force fields for molecular dynamics simulations." *Methods Mol Biol* **443**: 63-88. For additional reading on force fields, see Mackerell, A. D., Jr. (2004). "Empirical force fields for biological macromolecules: overview and issues." *J Comput Chem* **25**(13): 1584-604. Ponder, J. W. and D. A. Case (2003). "Force fields for protein simulations." *Adv Protein Chem* **66**: 27-85.

study macromolecules like AMBER (Cornell, Cieplak et al. 1995), OPLS (Jorgensen and Tiradorives 1988) and CHARMM (MacKerell, Bashford et al. 1998; MacKerell Jr, Brooks III et al. 1998).

In atomistic force fields, each atom is assigned a type and a partial charge. Force fields classify atoms not only by their element but also by their bond order and chemistry, hence its type. For example, aliphatic and aromatic carbon atoms each have different properties and are treated as different types. Another example would be primary, secondary, tertiary, and aromatic amines each of which provides a different numbers of hydrogen bond donors or acceptors. Force fields generally explicitly define all heavy atoms, however the treatment of protons differ. “All atom” force fields treat all protons explicitly, both polar and non-polar. “Unified” force fields treat the protons implicitly combining them with the heavy atom with which they are bond. Some unified force fields treat only polar protons explicitly, combining non-polar protons with their heavy atom binding partner. As with protons, some force fields treat lone pairs explicitly, though this is not common.

In addition to type, each atom is assigned a partial charge. Physically the distribution of charge in a molecule can be envisioned as massive positively charged nuclei in negative clouds of electrons. Computationally however, in molecular mechanics force fields, each atom is modeled at a single point in three-dimensional space, and therefore each body is assigned a value to represent its charge at that point combining nucleus and electron cloud. Unless the atom is an ion, each atom is assigned value from  $< 1$  to  $> -1$  corresponding to the distribution of elections around each nucleus, hence its partial charge. The sum of all partial charges is equal to the formal charge of the molecule. Partial charges can be assigned corresponding to atom type, but they are most often calculated using a variety of quantum mechanical models and statistical models. Different methods for determining partial charge can affect the outcome of simulations. (Mobley, Dumont et al. 2007)

The general form of the force field function is

$$E_{total} = E_{bonding} + E_{nonbonding} + E_{other}, \quad 1-4$$

where  $E_{total}$  is the total energy of a system defined using the force field, and  $E_{bonding}$  and  $E_{nonbonding}$  are the energetic contributions of the bonding and non-bonding forces. (Guvench and MacKerell 2008)  $E_{other}$  is an optional term representing other energetic contributions not described by the bonding and non-bonding forces, and frequently describes solvation forces in implicit solvent models. (Wang, Donini et al. 2001; Mackerell 2004; Guvench and MacKerell 2008) Each of the terms in equation 1-4 is the sum of their component energies, as described below. Each component energy is the sum of forces acting on each body in the system. Overall, the  $E_{total}$  in an atomistic force field is the potential energy for a given conformation of atoms.

The bonding energies in the force field are described as

$$E_{bonding} = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} K_\chi[1 + \cos(n\chi - \sigma)] \quad 1-5$$

where  $E_{bonding}$  is the sum of the sums of the energies from the bond lengths, bond angles, and dihedrals in the system. (Guvench and MacKerell 2008) The first two terms model the potential energy as bonds vibrate and pairs of connected bonds vibrate relative to each other as springs using Hook's Law. In the "bonds" sum in equation 1-5,  $b$  represents the bond length,  $b_0$  representing equilibrium bond length, and  $K_b$  representing the spring constant. In the "angles" sum,  $\theta$  represents the angle between the two bonds.  $\theta_0$  and  $K_\theta$  are the equilibrium angle and the spring constant respectively. The periodic change in potential energy as the dihedral bond rotates is modeled using a harmonic function. In the "dihedrals" sum,  $\chi$  is the angle of rotation around the central bond.  $n$ ,  $\sigma$ , and  $K_\chi$  define the frequency, wavelength and potential energy barrier height as the atoms on opposite sides of the dihedral bond overlap and separate during the rotation. Other common terms used by force fields to define the energy of bonding interactions include improper torsion, used to maintain chirality, and ring bending terms.

The nonbonding energies are calculated by:

$$E_{nonbonded} = \sum_{nonbonded\ pairs\ ij} \left( \epsilon_{ij} \left[ \left( \frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{min,ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon r_{ij}} \right) \quad 1-6$$

where  $E_{nonbonded}$  is the sum of energy of all pair-wise (i,j) non-bonding interactions between each body in the system as a function of distance,  $r_{ij}$ . (Guvench and MacKerell 2008) In equation 1-6 the two non-bonding interactions are defined. The first is the van der Waals (VdW) interactions using the classic 6-12 relationship.  $R_{min,ij}$  and  $\epsilon_{ij}$  are the constants defining the distance and depth of the energy minimum and are determined for each atom type pairing. The second interaction is charge-charge interactions modeled using Coulomb's Law.  $q_i$  and  $q_j$  are the partial charges on each body and  $\epsilon$  is the dielectric constant. Other non-bonding terms sometimes used in force fields include hydrogen bonds, modeled with an H-bonding 10-12 relationship, dipole, and multi-pole interactions.

#### 1.2.1.1.2 Force Fields in Docking

Force fields are employed to describe the binding energy landscape as potential energy function of the pose of the ligand. Docking programs such as DOCK 4.0 (Ewing, Makino et al. 2001), ADAM&EVE (Mizutani and Itai 2004), and EUDOC (Pang, Perola et al. 2001) use the AMBER force field for scoring. Theoretically, a good force field based on first principles should be sufficiently robust to simulate any system and be used for any calculation. In practice, force fields used to score docking, such as AMBER and CHARMM, were built for the simulation of macromolecules and are empirically parameterized using experimental data. (Guvench and MacKerell 2008) These force fields are designed to model the movement of amino acids in the target and not the small molecules of the ligand. Some force fields have been expanded to include generalized parameters for all organic compounds, such as Generalized AMBER Force Field, or GAFF. (Wang, Wolf et al. 2004) Additionally, while spectroscopic, thermodynamic, and crystallographic data is used for parameterization of the force fields,

binding data is not. (Guvench and MacKerell 2008) Thus, the direct use of force fields for scoring binding extends their use beyond their parameterization.

To overcome the limitations imposed by force field parameterization, docking programs use two approaches. In the first approach, programs such as Glide use the OPLS-AA force field only for minimization and MC, for which it was designed, but not final scoring. (Friesner, Banks et al. 2004) The second approach is to parameterize force fields using binding data. The scoring function is then the sum of the intermolecular interactions potential as defined in MM force fields with each term weighted based on binding parameterization. These scoring functions are known as empirical force field based scoring functions (EFFBSF). ROSETTA (Davis and Baker 2009), Gold (Jones, Willett et al. 1997) and AutoDock (Morris, Goodsell et al. 1998; Huey, Morris et al. 2007) use these scoring functions. The AutoDock scoring function has also been used in other docking programs like SODOCK (Chen, Liu et al. 2007) and ISE-Dock (Gorelik and Goldblum 2008), and will be described in more detail below.

Even with simplified EFFBSF, force fields are the sum of pair-wise interactions and can be computationally expensive. In order to decrease the computational cost of scoring, docking programs like AutoDock4 (Morris, Goodsell et al. 1998; Huey, Morris et al. 2007) and DOCK 4.0 (Ewing, Makino et al. 2001) use prebuilt scoring grids. Scoring grids are built by incrementally translating a probe atom through the predefined search volume on the target and scoring the probe at each point. A collection of grids are built for each atom type used by the scoring function which provides look-up tables for the docking program. However, the density of the point on the grid limits the approximation of the binding energy landscape by the scoring function.

#### ***1.2.1.2 Empirical Based Scoring Functions***

Empirical scoring functions are linear equations that describe the binding energy. Like QSAR, fitting the scoring function to experimental data determines the weight of each of the

energy terms. (Kitchen, Decornez et al. 2004; Mohan, Gibbs et al. 2005) The basic equation is a sum of the free energy terms. (Bohm 1994) Each of the weighted terms may describe a type of intermolecular interaction between the ligand and its target, a change in conformation, the environment of the ligand, or the target on binding. (Eldridge, Murray et al. 1997; Bohm 1998; Kramer, Rarey et al. 1999; Wang, Lai et al. 2002) These differ from EFFBSF in that they are not refinements of MM force fields, but are independently determined and often include a wider variety of terms. Common intermolecular interactions terms include: ionic interaction (Rarey, Kramer et al. 1996; Kramer, Rarey et al. 1999), hydrogen bonding (Rarey, Kramer et al. 1996; Eldridge, Murray et al. 1997; Bohm 1998; Kramer, Rarey et al. 1999; Wang, Lai et al. 2002), interactions with aromatic groups (Rarey, Kramer et al. 1996; Bohm 1998; Kramer, Rarey et al. 1999), lipophilic interactions (Rarey, Kramer et al. 1996; Eldridge, Murray et al. 1997; Bohm 1998; Kramer, Rarey et al. 1999), van der Waals interactions (Wang, Lai et al. 2002)<sup>3</sup> and interactions with metals (Eldridge, Murray et al. 1997). Other terms include desolvation (Bohm 1998; Wang, Lai et al. 2002) and ligand deformation (Wang, Lai et al. 2002). Entropy is modeled as a binding penalty for each rotamer in the ligand. (Rarey, Kramer et al. 1996; Eldridge, Murray et al. 1997; Bohm 1998; Kramer, Rarey et al. 1999) Each term is composed of the weight in the form of a free energy and a contribution function describing the degree to which each term is present or absent in a given ligand. (Bohm 1994) For example, the contributing functions for interactions are the pair-wise sums of the interactions between the ligand and target. These interactions are described spatially, based on the atomic coordinates, but unlike force fields, the atoms are not described in terms of interaction potential but in the types of the interactions they participate in. The contributing functions select which interactions are present based on the distance and angle between each atom and the concavity or convexity of

---

<sup>3</sup> LUDI uses only the repulsion portion, Bohm, H. J. (1998). "Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs." J Comput Aided Mol Des **12**(4): 309-23.

the binding site. (Eldridge, Murray et al. 1997; Bohm 1998; Kramer, Rarey et al. 1999; Wang, Lai et al. 2002) Contributing functions can be as simple as counting the number of rotatable bonds. (Rarey, Kramer et al. 1996; Eldridge, Murray et al. 1997; Bohm 1998; Kramer, Rarey et al. 1999) Because empirical scoring functions are linear equations, they contain a correction term, or y intercept. (Rarey, Kramer et al. 1996; Eldridge, Murray et al. 1997; Bohm 1998; Kramer, Rarey et al. 1999; Wang, Lai et al. 2002)

There are many empirical scoring functions, each with a different combination of terms, but they are generally derived for Bohm's original equation. (Kitchen, Decornez et al. 2004; Mohan, Gibbs et al. 2005) Examples of empirical scoring functions include LUDI (Bohm 1998), ChemScore (Eldridge, Murray et al. 1997), X-SCORE (Wang, Lai et al. 2002), and FlexX's scoring function (Rarey, Kramer et al. 1996; Kramer, Rarey et al. 1999). Some examples of docking programs that use empirical scoring functions include: GOLD which uses a ChemScore variant (Verdonk, Cole et al. 2003), Glide which uses ChemScore for a final scoring of the pose determined using force fields (Friesner, Banks et al. 2004; Halgren, Murphy et al. 2004), and Vina which uses an X-SCORE variant (Trott and Olson 2010).

Knowledge-based scoring functions are similar to empirical scoring functions in that they rely on a training set to determine the weight of different parameters. Only in this case, the training sets are the crystal structures and the scoring function is parameterized to reproduce the binding pose and not the free energy of binding. (Kitchen, Decornez et al. 2004) ROSETTA, in ROSETTALIGAND, uses a weighted knowledge-based scoring function. (Davis and Baker 2009) ROTA (Hartmann, Antes et al. 2009) was used originally for side-chain optimization, but was extended to protein ligand interaction using potentials of mean force (PMF) based on the ratio of the mean probability of the distance of two atom types from both a binder and a decoy set.

Ideal for docking, empirical scoring functions are quick to solve and specifically parameterized to predict the FEB, or pose. However, two factors limit them. First, as with all empirical approaches, the training set biases the scoring function to a region chemical space defined by the ligands and target systems used for parameterization. Secondly, there is uncertainty in which terms to include in the scoring function. Including too many terms may lead to over fitting the data, thereby decreasing the predictive power of the scoring function. Selecting too few terms may limit the robustness of the scoring function, as different systems are dependent on different terms.

### ***1.2.1.3 Consciences Scoring Methods***

Conscience scoring is the combining of different scoring functions. (Charifson, Corkery et al. 1999; Kitchen, Decornez et al. 2004) Examples include CScore, which combines existing docking scoring functions from DOCK, GOLD, and FlexX, and PMF scoring to rescore poses (Clark, Strizhev et al. 2002), and MOSFOM, which uses both energy score and contact scores during the docking process. (Li, Zhang et al. 2009) Conscience scoring methods try to utilize the best of each scoring functions, however they can amplify shared limitations. While individual scoring functions may outperform consciences scoring in specific systems, on average consciences scoring perform better over multiple systems. (Verdonk, Berdini et al. 2004)

### **1.2.2 Search Methods**

The goal of search methods is to find the global minimum of the binding energy landscape defined by the scoring function as quickly and efficiently as possible. (Campbell, Gold et al. 2003) Search methods can be classified into three general approaches: iterative, stochastic, and simulation based. (Kitchen, Decornez et al. 2004) The following sections described and evaluated each general approach on its ability to explore the binding energy landscape.

### ***1.2.2.1 Iterative Search Methods***

Iterative, or systematic, search methods attempt to map fully the binding energy landscape by enumerating and evaluating all degrees of freedom. (Kitchen, Decornez et al. 2004) For example, the EUDOC program translates a rigid ligand through the search volume in user defined steps, usually 1.0 to 0.5 Å. (Pang, Perola et al. 2001) At each point, the ligand is rotated along all three axis of rotation in 2 to 10 degree steps. (Pang, Perola et al. 2001) EUDOC calculates the potential energy of the system at each point by using the AMBER force field and returns the lowest energy pose as the docking result. (Pang, Perola et al. 2001) Ligand flexibility is modeled by individually docking each conformer.

The advantage of a systematic search is that it fully maps the binding energy landscape to within the resolution of the step size, thus increasing the probability of finding a true global minimum for the search volume. The challenge with iterative search methods is combinatorial explosion. (Kitchen, Decornez et al. 2004) An increase in search volume or ligand flexibility, or a decrease in search step size, will geometrically increase the number of scoring evaluations needed to fully explore the energy landscape, thereby increasing the time to identify the global minimum. All other search methods, including stochastic and simulation-based methods, are designed to overcome this problem of combinatorial explosion by finding the global minimum without defining the whole energy landscape.

One approach to avoid combinatorial explosion is to change the model of the ligand and the receptor from an atomic model. Changing the model is similar to changing the scoring function in that the approximation of the binding energy landscape is changed, but differs in that the search components are also changed. For example, DOCK describes the ligand and the negative space of the target's site as a set of spheres, defining the molecular surfaces of each. (Kuntz, Blaney et al. 1982; Ewing, Makino et al. 2001) DOCK geometrically matches the ligand spheres to the target site spheres. (Kuntz, Blaney et al. 1982; Ewing, Makino et al. 2001) Another example of a simplified model is ADAM&EVE, which models only h-bonding

interactions. (Mizutani and Itai 2004) Changing the model can only reduce the search time if the new model is simpler, but the model still needs to be robust. The sphere model used in DOCK is effective as it reduces the number of ligand poses that need to be iteratively checked by reducing the number of locations that the ligand can be placed. (Kuntz, Blaney et al. 1982; Ewing, Makino et al. 2001) However, the simpler model of ADAM&EVE is only useable in systems highly dependent on H-bonding.

Another approach is hierarchical systematic searches in which the granularity and focus of the search is refined over a series of searches. For example, Glide uses a hierarchical systematic search method (Friesner, Banks et al. 2004; Halgren, Murphy et al. 2004), which starts by performing an exhaustive enumeration of the ligand's rotamer space, selecting minimum energy conformations that are then screened in the binding site. (Friesner, Banks et al. 2004; Halgren, Murphy et al. 2004) An advantage of the initial ligand rotamer selection is that it is independent of the target and can be done once for any ligand and then reused for different targets. Hierarchical methods effectively reduce search time by making each search additive instead of multiplying the search with each new degree of freedom. However, each search partitions the binding energy landscape explored, which may lead the search away from the true global minimum.

First developed for *de novo* ligand design, incremental construction methods build a ligand by incrementally extending a base structure with new functional groups. (Jorgensen 2009) Incremental construction methods are a type of hierarchical systematic searches as the search is divided into steps, the most common implementation being root-branch methods. For example, FlexX is designed to model ligand flexibility by utilizing a root-branch incremental construction method. (Rarey, Kramer et al. 1996; Kramer, Rarey et al. 1999) FlexX divides a ligand into fragments at acyclic single bonds. The base, or root, fragment is docked into the target, and the original ligand is rebuilt by adding back the remaining fragments, allowing them to rotate on their connecting bonds. (Rarey, Kramer et al. 1996; Kramer, Rarey et al. 1999) Another example

is DOCK 4.0, which uses the space matching method described above to place the base fragment that is then extended to rebuild the ligand. (Ewing, Makino et al. 2001) Incremental construction methods suffer the same binding landscape partitioning limitations as hierarchical methods, but gain the advantage of decreased search times. Additionally, placement of the root can limit the accuracy of root-branch methods. The method assumes that the position root in the binding site corresponds to its lowest energy pose. This assumption is not always valid, as the lowest energy pose of the whole ligand might not correspond with the lowest energy pose of the root.

### ***1.2.2.2 Stochastic Search Methods***

Stochastic, or random, methods explore the binding energy landscape by making random perturbations and evaluating those perturbations to determine the next set of perturbations. (Kitchen, Decornez et al. 2004) The goal of random search methods is not to exhaustively explore the binding energy landscape like systematic methods, but to follow the surface of the binding energy landscape to the global minimum. Following the surface can significantly reduce the search time, as only the local area around each perturbation needs to be calculated. However, because the binding energy landscape is not completely explored, it is not known whether the minimum found is the true global minimum. Additionally, the found minimum are biased by the initial conditions, the number of scoring evaluations allowed (or steps taken), and the selection criteria for each step. To overcome biasing, random search method results are based on the consensus of multiple searches using different initial conditions.

Docking programs use a variety of random search methods including methods like Monte Carlo (MC), particle swarm optimization, and genetic or evolutionary methods. (Kitchen, Decornez et al. 2004) While MC methods are stochastic methods, they are also simulation methods and are discussed below. SODOCK uses particle swarm optimization to explore pose space using particles that are assigned vectors that describe a ligand pose based on the scoring function. (Chen, Liu et al. 2007) The vectors, treated as velocities, are accelerated using weighted random steps to explore local and global pose space. (Chen, Liu et al. 2007) The

optimum, lowest energy pose is determined by the consensus of neighboring particles in the swarm. (Chen, Liu et al. 2007) Genetic algorithms are used by docking programs like MolDock (Thomsen and Christensen 2006; De Azevedo 2010), ISE-Dock (Gorelik and Goldblum 2008), GOLD (Jones, Willett et al. 1997; Verdonk, Cole et al. 2003), and AutoDock3 and AutoDock4 (Morris, Goodsell et al. 1998; Morris, Huey et al. 2009). Genetic algorithms utilize the principles of natural selection to explore pose space. Each pose is defined as a gene to which random mutations are made, changing the pose. The fitness of the new gene is then evaluated based on a scoring function; the best gene becomes the foundation for the next generation.

All random search methods suffer the limitation of becoming stuck in local minima or not finding the global minimum as described above. However, due the dependence on censuses results, most docking programs using random search methods give a collection of poses for results as opposed to one global minimum pose. The multiple results can give insight to different binding poses that the ligand may adopt, as binding is dynamic.

### ***1.2.2.3 Simulation Based Search Methods***

Simulation-based search methods are based on the molecular simulation methods of molecular dynamics (MD), Monte Carlo (MC), molecular minimization, and simulated annealing. (Kitchen, Decornez et al. 2004) MD simulations model molecular and atomic motion using Newtonian physics. While MD simulations are extensively used in CADD (Alonso, Bliznyuk et al. 2006), its inability to cross high energy barriers make it a poor tool for fully exploring the binding energy, and is not commonly used as a search method. (Kitchen, Decornez et al. 2004) However, MC simulations can cross high-energy barriers, and therefore are used more commonly as a search method. MC is a stochastic method in which atom and molecules are randomly perturbed and the perturbations are accepted or rejected based on selection criteria, often an energy score. Grand Canonical MC simulations have been used to add ligands to a system. (Clark, Guarnieri et al. 2006; Clark, Meshkat et al. 2009) Minimization and simulated

annealing methods search for local minima in the binding energy landscape as determined by the gradient of the landscape's slope.

Some examples of docking programs that use simulation based search method are ROSETTALIGAND, Glide, AutoDock2 and Vina. ROSETTALIGAND uses MC and minimization to refine ligand and side-chain packing, utilizing rotamer libraries for the side-chain movement, followed by quasi-Newton minimization. (Meiler and Baker 2006; Davis and Baker 2009) Glide uses MC and minimization to refine its systematic search results. (Friesner, Banks et al. 2004; Halgren, Murphy et al. 2004) AutoDock2 uses a simulated annealing method to search for binding poses. (Goodsell and Olson 1990) Vina uses a gradient following minimization with multiple initial poses. (Trott and Olson 2010) Additionally, simulation-based search methods are commonly used to locally refine a pose as opposed to global searches because of their computational intensity.

#### ***1.2.2.4 Combining Search Methods***

Docking programs commonly combine multiple search methods to increase the granularity for the search, and to take advantage of the strengths of different methods. Many docking programs use local minimization to refine a pose identified by iterative or random search methods. Programs such as SODOCK (Chen, Liu et al. 2007), DOCK 4.0 (Ewing, Makino et al. 2001), and AutoDock4. (Morris, Goodsell et al. 1998; Huey, Morris et al. 2007; Morris, Huey et al. 2009), utilize this method. Another approach is hierarchical multistage searches. ROSETTALIGAND is an example of a multistage method beginning with a course-grain random search that is refined by using MC and then gradient minimization. (Davis and Baker 2009) Glide uses MC and then minimization to refine its systematic search results as well. (Friesner, Banks et al. 2004; Halgren, Murphy et al. 2004) Finally, consciences docking poses can be generated using multiple docking programs. ConsDock uses hierarchical clustering to select a consciences pose from poses generated by DOCK 4.0, FlexX, and Gold. (Paul and

Rognan 2002) Combining search methods can increase search speed by optimally using each search function's strengths, however, each search method used will bias the results.

### 1.2.3 Docking Performance

Docking programs combine search and scoring functions in a variety of ways, but, after many studies, no single program has significantly distinguished itself for its predictive qualities in all target systems (Bursulaya, Totrov et al. 2003; Kellenberger, Rodrigo et al. 2004; Kontoyianni, McClellan et al. 2004; Kontoyianni, Sokol et al. 2005; Chen, Lyne et al. 2006; Warren, Andrews et al. 2006; Cross, Thompson et al. 2009), nonetheless, some trends have appeared. For example, incremental build programs underperform when compared to other programs (Kellenberger, Rodrigo et al. 2004), and consciences scoring performs better on average (Verdonk, Berdini et al. 2004), yet the performance of a docking program is dependent on the target system. (Kellenberger, Rodrigo et al. 2004; Verdonk, Berdini et al. 2004; Warren, Andrews et al. 2006; Cross, Thompson et al. 2009) Generally, all docking programs share the same strengths and limitations. The studies show that docking programs are able to reproduce experimentally determined docking poses, but their scores do not correlate with experientially determined binding affinities. (Best shown in Warren et. al. (Warren, Andrews et al. 2006))

Docking programs can accurately, within a RMSD of 2 Å, reproduce experimentally determined binding poses (Kellenberger, Rodrigo et al. 2004; Chen, Lyne et al. 2006; Warren, Andrews et al. 2006; Cross, Thompson et al. 2009), however, the accuracy of pose prediction can be limited by both the ligand and the target. Pose prediction can be biased by the starting pose of a ligand. (Kellenberger, Rodrigo et al. 2004) Pose predictions started with experimentally determined binding poses are more accurate, by about 1 Å, than automatically generated or minimized poses because they start in the ideal binding conformation. (Cross, Thompson et al. 2009) Similarly, the conformation of the target can bias the docking results.

In addition, most modern docking programs treat ligands flexibly, either by allowing bond rotations during searching (e.g., AutoDock), or by rotating added fragments in incremental build docking (e.g. FlexX). However, ligand flexibility is generally limited to dihedral bond rotation. The more flexible the ligand and the functional groups are, the more degrees of freedom that need to be searched, thus decreasing the likelihood of finding the global minimum. (Chen, Lyne et al. 2006; Warren, Andrews et al. 2006) (Cross, Thompson et al. 2009)

While there is no correlation between the quality of the docking pose and the experimental resolution of the target structures (Chen, Lyne et al. 2006), the conformation of the receptor affects the quality of the docking pose. (Murray, Baxter et al. 1999) Most docking programs incorrectly treat the target as rigid. Different approaches have been applied to model receptor flexibility, though no single method is generally used. One approach is to allow for side-chain movement. AutoDock4 allows for selected side-chains to be treated flexibly. (Huey, Morris et al. 2007; Morris, Huey et al. 2009) ROSETTALIGNAD uses MC and minimization to allow side-chain movement and ligand repacking. (Davis and Baker 2009) Some programs, for example ADAM, “soften” the penalty of steric clashes by using an offset or soft interaction grid (Mizutani, Takamatsu et al. 2006). While allowing side-chain movement can improve docking pose, it is computationally more demanding and misses target conformational changes in the backbone. Multiple docking runs can be made against multiple static receptor conformations obtained experimentally or by modeling. (Totrov and Abagyan 2008) The challenge is selecting biologically relevant conformations and scoring the different target conformations relative to each other. The most common practice is to dock into the experimental structure that comes from a co-crystal, providing an open or bonded conformation for the target. In addition to flexibility, waters and cofactors that share the binding site can affect binding pose. (Verdonk, Chessari et al. 2005)

While the scoring functions are able to predict pose, they do not correlate with FEB (Warren, Andrews et al. 2006). This is likely because the scoring functions are evaluated on the

interactions of a single static pose between the ligand and the receptor, modeling enthalpic contributions to FEB, but not entropic contributions from either solvation or movement. Some scoring functions include terms to model both entropy and solvation, but often these estimates are inaccurate. For example, assigning a fixed entropic penalty to each rotatable bond, assuming they are fixed upon binding, overestimates the entropic penalty for ligands with more than two rotatable bonds. (Singh and Warshel 2010) FEB is an emergent property of the complex and the dynamic interactions between the ligand and its receptor, and therefore is most effectively modeled dynamically. (Mobley and Dill 2009)

#### 1.2.4 Example Program: AutoDock4

AutoDock4 combines a global random search function with a simulation-based local search, and scores with a semiempirical force field based scoring function. (Morris, Goodsell et al. 1998; Huey, Morris et al. 2007) AutoDock4's search method and scoring functions were originally developed for AutoDock3, but have been extended to allow for limited amino acid side-chain flexibility and a more robust scoring function. We used AutoDock4 for all our docking studies in this work. We selected AutoDock4 because it is freely available and widely used. Additionally, we selected it because it employs an empirical force field based scoring function. Most importantly, it was selected because of its satisfactory performance in testing. Using AutoDock4 provides a single program that can represent, in a limited way, a baseline for the current state of docking technology. This section provides a description of AutoDock4 to serve both as an introduction for future discussions and as an example of how a docking program's search and scoring functions work together.

AutoDock4 uses a semiempirical force field based scoring function. (Huey, Morris et al. 2007) The free energy of binding is defined by the scoring function as

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + (V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf}) \quad 1-7$$

where  $V$  is the sum of potential energy scores for each state's atomic pair-wise interactions and  $\Delta S_{conf}$  is the conformational entropy.  $L$  stands for ligand and  $P$  for protein. The L-L and P-P potentials model the intramolecular interactions, while P-L models the intermolecular interactions. The assumption is made that the distance between the ligand and the protein in the unbound state is sufficiently large enough for there to be no interaction, and therefore  $V_{unbound}^{P-L}$  is set to zero.  $\Delta S_{conf}$  is defined as

$$\Delta S_{conf} = W_{conf} N_{tors} \quad 1-8$$

where  $N_{tors}$  is the number of torsion angles and  $W_{conf}$  is the empirically derived weight of the term. Overall, the free energy of binding, defined by the scoring function, is the sum of the change in energy for both the change in ligand and protein conformation upon binding, the interaction energy between the ligand and the protein upon binding, and an entropy potential.

The equation for the interaction potentials is

$$V = W_{vdw} \sum_{ij} \left( \frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6} \right) + W_{H-bond} \sum_{ij} E(t) \left( \frac{C}{r_{ij}^{12}} - \frac{D}{r_{ij}^{10}} \right) + W_{elec} \sum_{ij} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} + W_{sol} \sum_{ij} (S_i V_j + S_j V_i) e^{-r_{ij}^2 / 2\sigma^2} \quad 1-9$$

where the terms are the sums of the energy contributions of each atom's pairs (i, j) in the system for the VdW interaction, hydrogen bond interactions, electrostatic interactions, and solvation, respectively. In all terms,  $r$  is the distance between each atom pair. The first term models VdW interactions using the standard 6-12 relationship. The AMBER force field provides the A and B constants for each atom pair type. The second term models H-bonding interactions using a 12-10 relationship.  $E(t)$  scales the 12-10 energy as the i, j angle varies from ideal. C and D determine the energy-well depth. The third term models electrostatic interactions using Coulombs' law as described in Section 1.2.1.1.1. The final term models the desolvation based on the volume of the atom,  $V$ , and the solvation parameter,  $S$ .  $\sigma$  is the weighted distance factor.  $W$  is the empirically determined weight of each term as determined by a training set of 288 protein ligand complexes,

of which 100 are HIV-1 protease complexes. To decrease computational time AutoGrid4 is used to build scoring grids for each atom type.

AutoDock4's search function combines a Lamarckian genetic algorithm for a global search with Solis-Wets minimization (Solis and Wets 1981) for local optimization. (Morris, Goodsell et al. 1998) As with all genetic algorithms, the pose of the ligand, or phenotype, is described as a series of genes. AutoDock4 uses three genes: (1) the Cartesian coordinates of the geometric center of the ligand in the same reference frame as the target, (2) quaternion to describe the orientation of the ligand, and (3) one gene describing each rotatable bond and its torsion angle. The search function begins by assigning random starting poses to each individual in a population. Each individual's fitness is evaluated using the scoring function described above, then a portion of the fittest from the population is minimized. The pose generated by the minimization is rescored and returned to the population. In this process, the phenotype becoming a new genotype is based on Lamarck's discredited evolutionary theory, hence the name Lamarckian genetic algorithm. After minimization and rescoring, a portion of the fittest members of the population becomes the basis for the next generation. The children are derived by randomly mutating each gene. As part of the child generation process, a number of crossover events are allowed to expand the search. At the same time, a number of elite poses are passed on to the next generation with no mutations to focus the search. The cycle of fitness selection with minimization continues until that population converges, having found a minimum, or after a user defined limit of energy calculations have been made (energy calculation being the most computational demanding step in the docking process). Due to the limitation of random search methods (biases inherent in the starting poses and the incomplete exploration for the binding energy landscape), AutoDock4 performs multiple searches using new starting populations. The results of each run are clustered based on pose to generate a list of scored possible poses.

### 1.3 VIRTUAL SCREENING

The goal of this work is to improve structure-based virtual screening. In general, virtual screening is the process of ranking a virtual chemical library according a metric. In structure based virtual screening, each member of a virtual chemical library is docked into a selected target generating a pose and a score. The scores form the metric that ranks the compounds with the assumption that better scoring compounds are more likely to bind to the target. This section begins with a discussion in Section 1.3.1 on virtual compound libraries in the context of virtual screening, looking at their sources, compound preparation, filtering, and library biases. This section concludes in Section 1.3.2 with an evaluation of the performance of virtual screening.

#### 1.3.1 Virtual Chemical Libraries

Chemical libraries are essential to the drug discovery process, as they are the source of lead compounds. (Zhou 2011) Their goal is to represent portions of chemical space, be that as diverse or focused as a drug discovery projects needs. Virtual chemical libraries, or compound databases, mimic their physical counterparts as collections of compounds sharing the same goals. There are a variety of methods ranging in complexity and information content used to represent compounds and their structures. (Zhou 2011) An example of a simple representation is SMILES (Simplified Molecular Input Line Entry Systems) which represent the chemical structures as a string using specific characters to represent bonds and groupings. (Weininger 1988; Weininger, Weininger et al. 1989; Weininger 1990) Examples of more complex representations are the .mol and .mol2 file types that have information on the coordinates of each atom in the structure, the bonds between the atoms and the order of each bond, and other relevant chemical information. In virtual screening, compound databases provide the library against which the target is screened.

Chemical libraries are obtained from a variety of sources including chemical vendors or public databases. Some commonly used libraries for virtual screening include the Cambridge structural database (Allen 2002), ZINC (Irwin and Shoichet 2005), NCI diversity set ([http://dtp.nci.nih.gov/docs/3d\\_database/Structural\\_information/structural\\_data.html](http://dtp.nci.nih.gov/docs/3d_database/Structural_information/structural_data.html)), and

PubChem (<http://pubchem.ncbi.nlm.nih.gov/>). Regardless of the compound database's source, each compound in the library needs to be compatible with the docking program selected for the virtual screening. Commonly, libraries are stored as SMILES or as 2D structures. Docking programs generally require 3D atomic coordinates for each compound. If not experimentally determined, structures can be generated automatically with programs like CORINA, CATALYST, OMEGA, and RUBICON. (Knox, Meegan et al. 2005) Compounds in the library need to be correctly protonated, depending on the pH of the target's environment or local conditions at the target site in order to generate accurate poses. (Knox, Meegan et al. 2005) Other compound preparations may be required for docking and virtual screening, for example the assignment of partial charges, indicating rotatable bonds, or building conformer libraries. A correctly prepared database is critical to an effective virtual screening, as incorrect preparation can generate bad poses and incorrect scoring. (Knox, Meegan et al. 2005) An example of a compound database designed specifically for virtual screening is the ZINC database. (Irwin and Shoichet 2005) ZINC is a collection of over 6 million commercially available compounds. (Irwin and Shoichet 2005) 3D coordinates are available for each compound for multiple protonation states with pre-assigned partial charges. (Irwin and Shoichet 2005) Compounds are stored as SMILES, mol2, 3D SDF, and DOCK flexibase formats. (Irwin and Shoichet 2005)

Chemical libraries are designed to be diverse, representing as much of chemical space as possible, or are focused on a specific subset of chemical space to obtain desired properties. (Zhou 2011) Compound databases can be focused by filtering for specific properties. Two common filters are "lead-like" and "drug-like". Drug-like compounds are filtered based on Lipinski's Rule of Five, which is based on a study of all drugs to select for optimum absorption. (Lipinski, Lombardo et al. 2001) Lipinski's Rule of Five is that poor absorption occurs where there are more than 5 H-bond donors, 10 H-bond acceptors, a molecular weight above 500 Daltons, and a calculated LogP (permeability) of greater than 5. (Lipinski, Lombardo et al. 2001) Because most compounds identified through virtual screening or high through-put screening are

likely to be synthetically modified, Opera developed the lead-like rules (e.g., 450 Daltons, CLogP between 4.5 and -3.5, no more than 4 rings, 10 non-terminal single bonds, 5 H-bond donors and 8 H-bond acceptors) based on known lead compounds. (Teague, Davis et al. 1999; Oprea, Davis et al. 2001; Oprea, Allu et al. 2007) Other filters may be target specific, such as only uncharged compounds, or only compounds with a specific functional group. Filtering a library, however, introduces bias by limiting the chemical space searched. (Verdonk, Berdini et al. 2004) Filtering can aid in a virtual screening by focusing the search to only compounds of interest, thereby reducing the computational time required for the screening. However, reducing the chemical space explored inhibits the identification of novel and potentially more potent leads. Ultimately, the degree to which a compound database is filtered is a balance between project requirements and computational resources.

### **1.3.2 Virtual Screening Performance**

Virtual screening has successfully identified many bioactive compounds, has led to the development of drugs (Villoutreix, Eudes et al. 2009; Talele, Khedkar et al. 2010), and has been extensively evaluated (Bursulaya, Totrov et al. 2003; Kellenberger, Rodrigo et al. 2004; Kontoyianni, McClellan et al. 2004; Kontoyianni, Sokol et al. 2005; Chen, Lyne et al. 2006; Warren, Andrews et al. 2006; Cross, Thompson et al. 2009). Virtual screening performance is based on a docking program's performance. As discussed in Section 1.2.3, docking programs can accurately reproduce experimentally determined binding poses (Kellenberger, Rodrigo et al. 2004; Chen, Lyne et al. 2006; Warren, Andrews et al. 2006; Cross, Thompson et al. 2009), but their scoring functions do not correlate with experimental binding data. (Warren, Andrews et al. 2006) Nevertheless, virtual screening can enrich compound databases for active compounds. (Bursulaya, Totrov et al. 2003; Kellenberger, Rodrigo et al. 2004; Kontoyianni, McClellan et al. 2004; Kontoyianni, Sokol et al. 2005; Chen, Lyne et al. 2006; Warren, Andrews et al. 2006; Cross, Thompson et al. 2009) In other words, the distribution of active compounds near the top of the rank order compound database is greater than a random distribution. Enrichment is plotted

with the percent database screened on the x-axis and the percent binders found on the y-axis. An ideal enrichment would be a curve shifted to the far left, while a random enrichment is the  $y=x$  line. The area under the curve (AUC) is a common numeric representation of the curve, with  $\sim 1$  as ideal and 0.5 as random. Mean AUC for virtual screening is between 0.55 and 0.72, therefore better than random. (Cross, Thompson et al. 2009)

Virtual screening suffers from two general problems. The first problem is the lack of robustness of the docking programs. Virtual screening performance is highly dependent on the target system and no docking program performs better than the rest. (Kellenberger, Rodrigo et al. 2004; Kontoyianni, McClellan et al. 2004; Warren, Andrews et al. 2006; Cross, Thompson et al. 2009) Additionally, virtual screening performance improves with expert involvement (Warren, Andrews et al. 2006; Cross, Thompson et al. 2009) The problem requires that each virtual screening project needs to first determine which docking program will work best for its system, and that in their current state of development docking programs cannot be used as a “black box”. The second major problem that virtual screening suffers is the large number of false positives it generates. (Bursulaya, Totrov et al. 2003; Kellenberger, Rodrigo et al. 2004; Kontoyianni, McClellan et al. 2004; Kontoyianni, Sokol et al. 2005; Chen, Lyne et al. 2006; Warren, Andrews et al. 2006; Cross, Thompson et al. 2009) While virtual screening can enrich a database, extensive testing is still required to find hits. A common occurrence is to have only a few percent of tested compounds show activity from virtual screening project. It is commonly thought that enrichment is dependent on the quality of pose prediction (Kellenberger, Rodrigo et al. 2004; Verdonk, Berdini et al. 2004), however, this has been shown not to always be the case (Warren, Andrews et al. 2006), suggesting that the main problem is the scoring function. The goal of this work is to improve virtual screening, thus it will focus on improving the scoring functions, as docking programs generally produce good poses. An improved scoring function is one that reduces the number of false positives and effectively discriminates between binders and non-binders in a variety of systems.

## 1.4 STATISTICAL MECHANICS BASED FREE ENERGY OF BINDING METHODS

Scoring functions determine the FEB based on a single static pose. With statistical mechanics (SM), the bulk of physical chemical properties, like FEB, are determined by the distribution of the microstates of the system. The equation for the FEB based on statistical mechanics is

$$\Delta G^\circ = -RT \ln \left( \frac{C^\circ \int e^{-((U(r_{PL})+W(r_{PL}))/RT)} dr_{PL}}{8\pi^2 \left( \int e^{-((U(r_P)+W(r_P))/RT)} dr_P \right) \left( \int e^{-((U(r_L)+W(r_L))/RT)} dr_L \right)} \right) \quad 1-10$$

where P is the target, L is the ligand, PL is the complex, R is the gas constant, T is the absolute temperature,  $C^\circ$  is standard concentrations, r is internal coordinates of a system, U(r) is potential energy as a function of the coordinates of conformations with W(r) as the corresponding solvation energy. (Gilson, Given et al. 1997; Gilson and Zhou 2007) A simplified form of this equation is

$$\Delta G^\circ = \langle U_{PL} \rangle - \langle U_P \rangle - \langle U_L \rangle + \langle W_{PL} \rangle - \langle W_P \rangle - \langle W_L \rangle - T\Delta S_{config}^\circ \quad 1-11$$

where  $\langle \rangle$  corresponds to the ensemble average or bulk property determined by the microstates. (Gilson and Zhou 2007) This equation defines FEB as an emergent property of the changing interaction potentials of the ligand and the target as they move relative to each other and their respective environment. Therefore SM based FEB calculations may resolve the challenges present in virtual screening.

The microstates of systems can be determined via molecular mechanics methods (e.g., MD, MC), with the potential energies of each microstate determined by the force field. While sampling all microstates is impossible, is it possible to sample a sufficient number of states to estimate accurately the FEB. The ideal molecular FEB simulation would be to simulate the ligand and target in an explicate solvent, allowing sufficient runtime for the ligand and the target to associate and disassociate multiple times. Currently, atomistic versions of the ideal simulation are virtually impossible, therefore a variety of methods are used to simplify the ideal

simulation, sampling only the key stages in the binding process. These methods are categorized into two types: end point methods, and pathway methods.

The goal of this section is to introduce the two different statistical mechanics computational approaches used to calculate the FEB using molecular mechanics methods (Sections 1.4.1, and 1.4.2), and examine their application to drug discovery (Section 1.4.3). Section 1.4.4 concludes with an introduction to the mean field free energy of binding (MF-FEB) method for calculating FEB.

### 1.4.1 End Point Methods

Because FEB is a state function, end point methods calculate FEB based on the difference in energy between the bound and unbound state.(Alonso, Bliznyuk et al. 2006; Gilson and Zhou 2007) The energies of the states are determined using molecular mechanics approaches (i.e., MD simulations). The two most commonly used end point methods are linear interaction energy (LIE), and the molecular mechanics and Poisson-Boltzmann surface area (MM-PBSA). Each method is reviewed below.

#### 1.4.1.1 LIE

Originally developed by Aqvist et. al., the linear interaction energy, or LIE, method calculates the absolute free energy of binding based on an empirically solved linear equation. The equation is a function of the differences in average potential energy between MD or MC generated ensembles of the ligand bound to the target, and the ligand unbound. (Aqvist, Medina et al. 1994; Aqvist and Marelius 2001) The basic equation is

$$\Delta G_{\text{Binding}} = \alpha \langle V_{\text{bound}}^{\text{elec}} - V_{\text{unbound}}^{\text{elec}} \rangle + \beta \langle V_{\text{bound}}^{\text{vdw}} - V_{\text{unbound}}^{\text{vdw}} \rangle + \gamma \quad 1-12$$

where  $\langle V-V \rangle$  is the difference in the ensemble average interaction potentials between the bound and unbound state based on the force field used to simulate the system.(Alonso, Bliznyuk et al. 2006)  $\alpha$ ,  $\beta$ , and  $\gamma$  correspond to the slopes and intercept of the linear equation which are determined uniquely for each systems.  $\alpha$  is about 0.5 corresponding to the first order

approximation for the electrostatic contribution to free energy of binding, but varies with the system. (Aqvist, Medina et al. 1994; Wang, Wang et al. 1999)  $\beta$  is unique for each system and correlates with non-polar desolvation. (Wang, Wang et al. 1999)  $\gamma$  was not included in the original development of LIE to avoid over fitting (Aqvist and Marelus 2001), but was first added to account for solvent accessible surface areas (Smith, Jorgensen et al. 1998) and thus became a standard term, although it can go to 0 in some systems. (Aqvist and Marelus 2001; Alonso, Bliznyuk et al. 2006)

When compared to the SM-FEB equation, LIE utilizes its fitting to determine the energetic contributions from solvation, entropy, and ligand-target interactions not determined from the simulations. Therefore, LIE is often of limited use in drug discovery because  $\alpha$ ,  $\beta$ , and  $\gamma$  are determined empirically for each system requiring training sets of known binders (ideally with experimentally determined poses). LIE cannot be used on binders for a “new” system in the early stages of drug discovery, but LIE is useful for the refinement of compounds once hits have been discovered, as long as the new compounds do not deviate from the chemical space defined by the training set. When properly implemented, LIE performs well, predicting energies that correlate closely ( $r^2 > 0.7$ ) with experimental FEB. LIE has been used to find bioactive compounds in many different targets. Some more recent examples include adenosine deaminase (Kosugi, Nakanishi et al. 2009), tubulin (Alam and Naik 2009), HIV-1 reverse transcriptase (Carlsson, Boukharta et al. 2008; Nervall, Hanspers et al. 2008), neuraminidase (Park and Jo 2010), and MurD ligase (Perdih, Bren et al. 2009).

#### ***1.4.1.2 MM-PBSA***

Originally developed by Kollman and his coworkers (Kollman, Massova et al. 2000), the molecular mechanics and Poisson-Boltzmann surface area (MM-PBSA) is an end point method that determines the free energy of binding based on either a single MD simulation of a ligand bound to its target, or three MD simulations: one of the complex, one of the ligand, and one of

the target. (Alonso, Bliznyuk et al. 2006; Gilson and Zhou 2007) The overall equation to calculate the FEB using MM-PBSA is

$$\Delta G_{binding} = \bar{G}_{complex} - [\bar{G}_{protein} + \bar{G}_{ligand}] \quad 1-13$$

based on an end point approach. (Alonso, Bliznyuk et al. 2006) The equation for the average free energy contribution of each state is

$$\bar{G} = \bar{E}_{MM} + \bar{G}_{solvation} - T\bar{S} \quad 1-14$$

where  $E_{MM}$  is the average molecular mechanics energy from the simulation,  $G_{solvation}$  is the free energy of solvation, and  $TS$  is the entropy term. (Kollman, Massova et al. 2000) When only the ligand-target complex is simulated, the protein and ligand energy contribution are determined by reanalyzing the MD trajectories with the protein or ligand removed. Average molecular mechanics energy is determined by

$$\bar{E}_{MM} = \bar{E}_{bond} + \bar{E}_{angle} + \bar{E}_{tors} + \bar{E}_{vdw} + \bar{E}_{ele} \quad 1-15$$

or the sum of the average contributions from energetic components of the molecular mechanics force field with no cutoff range used in the MD simulations. (Kollman, Massova et al. 2000) The free energy of solvation is determined using a numerical solution for Poisson-Boltzmann equations for the electrostatic contributions and an estimate of the solvent-accessible surface area for the non-polar contributions. (Kollman, Massova et al. 2000; Alonso, Bliznyuk et al. 2006) The entropic contribution is determined by using quasi-harmonic or normal mode analysis of the MD trajectory. (Kollman, Massova et al. 2000; Alonso, Bliznyuk et al. 2006) A number of modifications exist for the MM-PBSA method, the most common being to replace the Poisson-Boltzmann equations with a generalized Born model. Some examples of other modification from recent works include the explicit and implicit solvation models (Jiao, Zhang et al. 2009), the addition polarizable force fields (Jiao, Zhang et al. 2009), and starting MD runs from a minimized system with no equilibration (Rastelli, Del Rio et al. 2010).

Equation 1-13 is identical in form to equation 1-11, therefore limitations in the MM-PBSA method comes from the ability to accurately compute each component from the MD simulations. Because entropy cannot be determined directly from the MD simulations, it is calculated separately, which can be computationally demanding and inaccurate. (Brown and Muchmore 2006) Protein solvation calculations have been shown to need improvement. (Guimaraes and Mathiowetz 2010) Convergence may be difficult due to energy contributions from changes in the whole protein structure. (Gilson and Zhou 2007) Despite some limitations researchers have successfully employed MM-PBSA calculations in a number of systems, a few examples include HIV protease (Chen, Yang et al. 2009; Chen, Zhang et al. 2009; Das, Koh et al. 2009), trypsin (Jiao, Zhang et al. 2009), RNA-protein binding (Kollman, Massova et al. 2000), and plasmepsin 2 (Degliesposti, Kasam et al. 2009). Additionally, in small developmental studies, the MM-PBSA methods can approximate FEB to about 1 to 2 kcal mol<sup>-1</sup>. (Alonso, Bliznyuk et al. 2006; Degliesposti, Kasam et al. 2009; Jiao, Zhang et al. 2009)

#### 1.4.2 Pathway Methods – Free Energy Perturbation

FEB of binding can be defined as:

$$\Delta G_{binding} = -RT \ln \langle e^{-\Delta U_{AB}/RT} \rangle_{...A} \quad 1-16$$

where  $\Delta U$  is the change in energy between two states (A and B), bound and unbound, and  $\langle \rangle$  is the ensemble average over the initial state. (Gilson and Zhou 2007; Cossins, Foucher et al. 2009) While the ensemble average can be determined from a single MD/MC simulation (i.e., our ideal simulation), unless the initial and final states are similar, convergence is difficult. (Gilson and Zhou 2007) To solve the convergence problem, intermediate steps are added thereby creating a path between the initial and final states. (Alonso, Bliznyuk et al. 2006; Gilson and Zhou 2007; Cossins, Foucher et al. 2009) The perturbation along series of states is governed using a coupling parameter,  $\lambda$ , that is varied between 0 and 1. (Alonso, Bliznyuk et al. 2006; Gilson and Zhou 2007; Cossins, Foucher et al. 2009) For each  $\lambda$  step, the system can then be simulated and the  $\langle e^{-\delta U/RT} \rangle$  determined. (Gilson and Zhou 2007) The FEB is calculated from the combination

of perturbations steps. This method is referred to as free energy perturbation (FEP). FEB is also calculated using thermodynamic integration (TI). TI integrates first derivatives of FEB that are determined from MD simulation over a path like FEP.

The implementations of FEP vary more than the implementations of LIE or MM-PBSA, but can be divided roughly into four groups in two families. The first family is relative FEB methods, with the most common method being alchemical FEP. (Gilson and Zhou 2007; Deng and Roux 2009; Mobley and Dill 2009) The relative FEB can be determined by perturbing or mutating one ligand into another. These alchemical pathways require that both ligands differ by only a functional group. (Helms and Wade 1998) The second family of methods is absolute FEB methods where the ligand is removed from the target using different pathways. (Gilson and Zhou 2007; Deng and Roux 2009; Mobley and Dill 2009) In an annihilation method, the simulation disappears or removes the ligand from the binding site over the perturbation path. (Jorgensen, Buckner et al. 1988) By using the potential of mean force, the ligand is incrementally pulled away from the target. (Woo and Roux 2005) The most common method is decoupling where the intermolecular interactions are turned off over the course of the perturbation. (Boresch, Tettinger et al. 2003; Deng and Roux 2006; Mobley, Chodera et al. 2007)

In general, FEP methods can predict the FEB of ligands with errors ranging for 1 to 3 kcal mol<sup>-1</sup> for decoupling methods. (Deng and Roux 2006; Mobley and Dill 2009; Michel and Essex 2010) In a small test set, FEP decoupling methods have been shown to outperform DOCK scoring. (Mobley, Graves et al. 2007) Additionally, in limited comparative testing, perturbation methods were shown to perform better than MM-PBSA. (Jiao, Zhang et al. 2009) However, PMF methods can still have difficulty converging and can be dependent on force field parameterization. (Alonso, Bliznyuk et al. 2006; Deng and Roux 2009; Michel and Essex 2010) FEP's main limitation is the need for extensive sampling which is computationally expensive. (Alonso, Bliznyuk et al. 2006; Michel and Essex 2010)

### 1.4.3 Application in Drug Discovery

Because the MD and MC simulations are computationally expensive, statistical mechanics based FEB methods are not widely used in drug discovery alone, but are used to post-process docking results. (Alonso, Bliznyuk et al. 2006) By using LIE to rescore docking generated poses, predicted FEB can be correlated with experimental results. (Carlsson, Boukharta et al. 2008) MM-PBSA methods have been used to rescore docking results. (Brown and Muchmore 2006; Guimaraes and Cardozo 2008; Thompson, Humblet et al. 2008) For example, rescoring FlexX with the BEAR program that included MM-PBSA and MM-GBSA showed retrospectively significant enrichment of known binders from the rescoring of the top 5000 hits of a one million compound database. (Degliesposti, Kasam et al. 2009) However, while considered the most accurate and rigorous approach to calculate FEB, FEP methods have not been used to rescore docking outside of very small test sets. (Mobley, Graves et al. 2007)

### 1.4.4 Mean Field Free Energy of Binding

Roux and coworkers developed the mean field method of calculating the absolute free energy of binding (MF-FEB) between a ligand and a protein. (Woo and Roux 2005; Deng and Roux 2006; Deng and Roux 2008) In summary, MF-FEB is a double decoupling FEP method. The binding free energy is calculated as the difference between the free energy of decoupling the ligand from its binding site and the free energy of decoupling the ligand from the solvent. Additionally, conformational and positional constraints that are later removed are placed on the ligand in the site. The removal of the constraints on the ligand models the ligand regaining entropic freedom as it leaves the binding site. This mean field approach aids in convergence by limiting the microstates that need to be sampled as the ligand is being decoupled because the ligand movement is constrained.

MF-FEB is based on

$$K_b = \frac{\int_{site} d(L) \int d(X) e^{-\beta U}}{\int_{bulk} d(L) \delta(r_L - r^*) \int d(X) e^{-\beta U}} \quad 1-17$$

where  $K_b$  is the binding constant, and the integrals are conformational integrals based on the coordinates of the ligand(L) or the target and solvent (X). The  $r$  in the equation represents the movement of the ligand from the binding site ( $r_L$ ) to an arbitrary location away from the target and in the bulk solvent ( $r^*$ ). Additional conformational integrals may be added as long as the ratio of those integrals is one, such that

$$\begin{aligned}
K_b = & \frac{\int_{\text{site}} d(L) \int d(X) e^{-\beta U_1}}{\int_{\text{bulk}} d(L) \int d(X) e^{-\beta[U_1+u_c]}} \times \frac{\int_{\text{site}} d(L) \int d(X) e^{-\beta[U_1+u_c]}}{\int_{\text{bulk}} d(L) \int d(X) e^{-\beta[U_1+u_c+u_t]}} \\
& \times \frac{\int_{\text{site}} d(L) \int d(X) e^{-\beta[U_1+u_c+u_t]}}{\int_{\text{bulk}} d(L) \int d(X) e^{-\beta[U_1+u_c+u_t+u_r]}} \times \frac{\int_{\text{site}} d(L) \int d(X) e^{-\beta[U_1+u_c+u_t+u_r]}}{\int_{\text{bulk}} d(L) \int d(X) e^{-\beta[U_0+u_c+u_t+u_r]}} \\
& \times \frac{\int_{\text{site}} d(L) \int d(X) e^{-\beta[U_0+u_c+u_t+u_r]}}{\int_{\text{bulk}} d(L) \delta(r_L - r^*) \int d(X) e^{-\beta[U_0+u_c+u_t]}} \\
& \times \frac{\int_{\text{site}} d(L) \int d(X) e^{-\beta[U_0+u_c+u_t]}}{\int_{\text{bulk}} d(L) \delta(r_L - r^*) \int d(X) e^{-\beta[U_0+u_c]}} \times \frac{\int_{\text{site}} d(L) \int d(X) e^{-\beta[U_0+u_c]}}{\int_{\text{bulk}} d(L) \delta(r_L - r^*) \int d(X) e^{-\beta[U_1+u_c]}} \\
& \times \frac{\int_{\text{site}} d(L) \int d(X) e^{-\beta[U_1+u_c]}}{\int_{\text{bulk}} d(L) \delta(r_L - r^*) \int d(X) e^{-\beta[U_1]}}
\end{aligned} \tag{1-18}$$

where  $u_c$  is the potential energy from the restraint on the conformation of the ligand;  $u_t$  and  $u_r$  are the potential energy from the translation and rotational constraints, respectively, from the positional constraint on the ligand as defined by a line connecting three atoms in the ligand to three atoms in the site;  $U_1$  is the potential energy corresponding to the ligand in an interaction with the site or the bulk (as in  $\lambda = 1$ ); and  $U_0$  is the potential energy corresponding to the ligand decoupled from the site or the bulk (as in  $\lambda = 0$ ). Each added conformational integral ratio adds a step in the process of confining and releasing the ligand as it moves from the site to the bulk. The steps are: (1) applying the conformation constraint to the ligand in the site, (2) applying the translation constraint to the ligand in the site, (3) applying the rotational constraint to the ligand in the site, (4) decoupling the ligand from the site with the constraints on, (5) removing the rotational constraint from the ligand, (6) removing the translation constraint from the ligand, (7) coupling the ligand into the bulk, and (8) removing the conformation constraint from the ligand.

Most of the conformational integrals can be expressed as ensemble averages as

$$K_b = \langle e^{-\beta u_c} \rangle_{(site; U_1)} + \langle e^{-\beta u_r} \rangle_{(site; U_1+u_c)} + \langle e^{-\beta u_r} \rangle_{(site; U_1+u_c+u_t)} + \langle e^{-\beta[U_1-U_2]} \rangle_{(site; U_1+u_c+u_t+u_r)} \\ + F_r + F_t + \langle e^{-\beta[U_1-U_2]} \rangle_{(bulk; U_1+u_c)} + \langle e^{-\beta u_c} \rangle_{(bulk; U_1)} \quad 1-19$$

where

$$F_t = \frac{\int_{site} d(L) \int d(X) e^{-\beta[U_0+u_c+u_t]}}{\int_{bulk} d(L) \delta(r_L - r^*) \int d(X) e^{-\beta[U_0+u_c]}} \quad 1-20$$

and

$$F_r = \frac{\int_{site} d(L) \int d(X) e^{-\beta[U_0+u_c+u_t+u_r]}}{\int_{bulk} d(L) \delta(r_L - r^*) \int d(X) e^{-\beta[U_0+u_c+u_t]}} \quad 1-21$$

where  $F_t$  and  $F_r$  are the translation and rotational factors.  $F_t$  and  $F_r$  can be solved directly with numeric integration schemes by simplifying them to where

$$F_t = \int dr_1 e^{-\beta u_t(r_1)} \quad 1-22$$

and:

$$F_r = \frac{\int d\Omega_1 e^{-\beta u_r(\Omega_1)}}{\int d\Omega_1} \quad 1-23$$

where  $\Omega_1$  is the set of three angles for rigid body rotation, and  $u_r(\Omega_1)$  is the restraint potential. MF-FEB calculations are independent of restraining potentials values. (Deng and Roux 2006)

By definition

$$K_b^o \stackrel{def}{=} e^{(-\beta \Delta G_{binding}^o)} \quad 1-24$$

And  $\Delta G_{binding}^o$  can be expressed as

$$\Delta G_{binding}^o = \Delta \Delta G_{int} + \Delta \Delta G_t^o + \Delta \Delta G_r + \Delta \Delta G_c \quad 1-25$$

where

$$\Delta \Delta G_{int} = (\Delta G_{int}^{site} - \Delta G_{int}^{bulk}) \quad 1-26$$

where  $\Delta\Delta G_{int}$  is the free energy of moving the ligand from the site to the bulk where  $\Delta G_{int}$  are the interaction energy between the ligand and the two systems

$$\Delta\Delta G_t^o = (-\Delta G_t^{site} - k_B T \ln(F_t C^o)) \quad 1-27$$

and

$$\Delta\Delta G_r = (-\Delta G_r^{site} - k_B T \ln(F_r)) \quad 1-28$$

corresponds to the free energy costs of the positional restraints, and

$$\Delta\Delta G_c = (\Delta G_c^{site} - \Delta G_c^{bulk}) \quad 1-29$$

where

$$\Delta G_c^a = -k_B T \ln \frac{\int d\zeta \rho_a(\zeta) e^{-\beta u_c}}{\int d\zeta \rho_a(\zeta)} \quad 1-30$$

where  $\rho_a(\zeta)$  is the unbiased distribution of the RMSD of the ligand where “a” is site or bulk is the free energy of cost of the conformational restraints.

MD based FEP methods are used to compute the ensemble averages. The site is modeled as a 15 Å radius solvent drop sphere centered on the ligand in the target site using a generalized boundary potential to model the solvent and protein beyond the sphere. The bulk is also modeled as a 15 Å radius solvent drop sphere centered on the ligand with explicit waters in the sphere and a solvent shell boundary potential. The interaction energy between the ligand and the respected systems is computed by the decoupling, the turning off, the electrostatic interactions, and the VdW dispersion and VdW repulsions, separated using the Weeks-Chandler separation (Weeks, Chandler et al. 1971). The constraint potentials are removed by perturbation in the site as outlined above. The energy of each step of the perturbation is calculated using a weighted histogram analysis. (Kumar, Rosenberg et al. 1992) By using these methods, the average reported error for FEB calculations was  $\sim 1$  kcal mol<sup>-1</sup> for eleven hydrophobic binders of L99A T4 Lysozyme (Deng and Roux 2006), and eight FK506 binders. (Wang, Deng et al. 2006)

Additionally, one non-binder test within T4 Lysozyme had a FEB greater than  $-1 \text{ kcal mol}^{-1}$ . (Deng and Roux 2006)

## **1.5 INTRODUCTION TO HIGH PERFORMANCE COMPUTING**

Predicting the properties of chemical compounds requires significant computational resources. These requirements can be due to repeating the same computation many times on different systems, like in virtual screening, or due to a large computation carried out on one system, like in MD simulations. High performance computers provide the required computational resources to predict the properties of chemical compounds.

In this work, we employ two high performance computing architectures: supercomputers, and grid, or distributed, computing. Section 1.5.1 introduces both architectures. Section 1.5.2 discusses the use of grid computing in drug discovery.

### **1.5.1 Types of High Performance Computers**

Supercomputers are large computers housed in a single location. Modern supercomputers are composed of nodes, which are a group of processors that share the same RAM, not unlike a single desktop computer. High-speed networks connect each node together allowing all processors to work in parallel. The nodes also share a common file system. The supercomputer's organization allows calculations, also called jobs, to be made on one or more nodes at a time. As the number of nodes increases the overall processing power increases, however, runtime is ultimately dependent on the bandwidth of the network connecting the nodes. The scalability of processing power makes supercomputers ideal for both single large jobs and similarly repeated smaller jobs.

Grid computers are composites of a number of different member devices, usually dispersed geographically, but controlled by a central server creating a single virtual machine. Unlike supercomputers, the member devices of a grid computer can have different processors and even operating systems. Each member device also works independently, getting its jobs

from the central server, often over the internet. The independence of the member device means that large-scale calculations that take advantage of the parallel architecture of supercomputers do not work well on a grid computer. However, grid computers are ideal for smaller jobs that can be run on a single processor. Because grid computers are easily expanded by just adding new member devices, the amount of computation a grid computer can do is often more than a common supercomputer. (Tomlinson, Malmstrom et al. 2009)

### **1.5.2 Drug Discovery with Grid Computing**

Grid computing has been effectively used in drug discovery, especially in highly parallelizable virtual screening projects. Researchers employ two types of grid architecture, public and private. Private grids are small in-house grids composed of the computers in a single institution. Brown and coworkers, for example, built a small private grid at their company to rescore docking generated poses using MM-PBSA. (Brown and Muchmore 2006) Hydra is an example of a program specifically developed to run virtual screening on a small grid. (Bullard, Gobbi et al. 2008) Private grids can also extend between academic and government institutions, for example the EGEE grid, which is composed of supercomputers across Europe. (Jacq, Salzemann et al. 2006; Jacq 2007) A major project of the EGEE grid is WISDOM (Wide In-Silico Docking On Malaria) that has successfully identified potential anti-malarial drugs through virtual screening and limited use of MM-PBSA. (Kasam, Zimmermann et al. 2007; Salzemann, Kasam et al. 2007; Degliesposti, Kasam et al. 2009; Kasam, Salzemann et al. 2009)

Private computer users have donated CPU time on their devices to form public grids. The member devices receive work from a grid server. The work is then run on the member devices while the computer is not being used thereby “harvesting” unused CPU cycles. BOINC is an example of a program designed to set-up and run public grids. (boinc.berkeley.edu) The Lifesaver Screensaver was an early example of a public grid virtual screening project. (Richards 2002) Later, IBM developed World Community Grid, which IBM provides as a service to researchers. Virtual screening projects on World Community Grid include Fight AIDS@Home

(Chang, Lindstrom et al. 2007), and Discovering Dengue Drugs-Together (Tomlinson, Malmstrom et al. 2009). World Community Grid is described in Chapter 5.

Grid computing provides significant computational resources and has been successfully used in drug discovery projects. However, running a grid requires significant resources in both time and equipment. Regardless of the grid's architecture, a grid requires a server to handle work distribution and collection. Additionally, grids have to be set-up and maintained, which may require a researcher to spend time as an IT specialties in addition to being a scientist. Public grids add the additional challenge of porting programs to different operating systems. The most successful grid drug discovery projects, like WISDOM and Fight AIDS@Home, happen when the grid is maintained by a separate organization from that doing the research, thereby utilizing the expertise of IT specialists to maintain the grid and allowing the scientists to focus on the research.

## **1.6 OVERVIEW OF WORK**

In the previous sections, we discussed the drug discovery and development process and how computers are used to predict chemical properties, particularly binding. We examined docking programs showing how their search and scoring functions can accurately predict experimental binding poses, but fail to predict accurately experimental free energy of binding. We showed that while large-scale docking studies of compound databases (virtual screening) enriches databases for potential binders, they also generate large numbers of false positives due to the inaccuracy of the scoring functions. In addition to scoring functions, we reviewed statistical mechanics methods of calculating the free energy of binding. Statistical mechanics FEB methods were shown to be more accurate at reproducing experimental free energy of binding. However, the computational demands of these methods are too great to use in conjunction with pose and conformation search methods requiring predetermination of a ligand's pose. The last section concluded with a brief introduction to high performance computing.

The goal for work is to improve drug discovery tools, specifically virtual screening. Our hypothesis is that enrichment in virtual screening can be improved over traditional scoring functions by rescoring docking generated poses with perturbation based statistical mechanics FEB calculations. This combined docking and statistical mechanics FEB rescoring uses both methods optimally with the docking programs generating the ligand poses, and perturbation statistical mechanics FEB calculations rescoring the poses with, ideally, increased accuracy. In order to make large-scale perturbation based FEB calculations, we employ the perturbation method of MF-FEB as described in the Section 1.4.4. The MF-FEB method is ideal for grid and supercomputing environments for two reasons. First, each step in the perturbation process is an independent simulation allowing them to be run in parallel. Second, the small size of the simulated system makes each job tractable on a common single processor device. To perform larger scale virtual screening with MF-FEB rescoring, we used the computing power of both Texas Advanced Computing Center (TACC) and IBM's World Community Grid.

The balance of this work discusses tests and applications of virtual screening with MF-FEB rescoring. Chapter 2 describes the methods used to perform virtual screening with MF-FEB rescoring using TACC's supercomputers. Chapter 3 examines the optimization of docking and MF-FEB calculations, and the filtering of virtual screening results. Chapter 4 covers a retrospective study of MF-FEB rescoring of AutoDock4 generated poses in a small test of known binders and non-binders to L99A T4 lysozyme. This chapter shows that docking generated pose can be used for MF-FEB calculations, and MF-FEB rescoring can discriminate between binders and non-binders better than AutoDock4's scoring function. Chapter 5 discusses the application of the virtual screening with MF-FEB rescoring in two projects run on IBM's World Community Grid, Discovering Dengue Drugs-Together (DDDT) and Influenza Anti-viral Drug Search (IADS). Chapter 6 describes the methods employed for virtual screening and MF-FEB calculation on World Community Grid. Chapter 7 examines a second retrospective study of virtual screening with MF-FEB rescoring in two systems taken from the Database of Useful

Decoys (Huang, Shoichet et al. 2006), showing that while MF-FEB rescoring improves enrichment, ultimately, MF-FEB rescoring suffers from similar problems as scoring functions. Chapter 8 reviews the finding of this work and describes further work that can be derived from it.

## Chapter 2 Methods for Docking, Virtual Screening, and Mean Field–Free Energy of Binding

This chapter describes the methods for docking, virtual screening, and MF-FEB calculations using the high performance computers Ranger and Lonestar at the Texas Advanced Computing Center (TACC) to test if MF-FEB scoring will improve enrichment over AutoDock4's scoring function. Section 2.1 lists the software and hardware used in this work. Section 2.2 describes the standard method for preparing ligand and target files. Section 2.3 describes docking and virtual screening with AutoDock4. The chapter concludes with Section 2.4 describing the methods used for MF-FEB calculations.

This chapter uses two naming conventions for convenience and clarity. First, in normal type, “ligand” refers to the small molecules used in binding simulations and calculations, and “target” refers the macromolecule the ligand is bound to. In all cases, the target is a protein, even though the type of protein may be an enzyme, a receptor, etc. Second, the italicized terms in file names are variable name components that are modified according to the work performed, yet are consistent in their usage. For example, *ligand* is the location of ligand name or identifier in a file name.

### 2.1 COMPUTATIONAL RESOURCES

This section provides a description of the software and hardware used in this work, and introduces the .pdb and .mol2 file formats.

#### 2.1.1 Software

This section described the software utilized in this work, and is subdivided by function.

##### 2.1.1.1 Docking

We performed all docking simulations using the AutoDock4 package that contains both the AutoDock4 and AutoGrid4 programs. (Huey, Morris et al. 2007; Morris, Huey et al. 2009) AutoDock4 is available from [autodock.scripps.edu](http://autodock.scripps.edu) under the GNU license. During the course of

this research, AutoDock4 was under development and was updated from version 4.01 to 4.2. Version 4.2 was the default version used in this work. We also used AutoGrid4 version 4.01, which was not updated. Additionally, AutoDock Tools (ADT) (Morris, Huey et al. 2009), also obtained from autodock.scripps.edu, was used to prepare docking files.

### **2.1.1.2 Molecular Simulation**

We used CHARMM c34a2 package (Brooks, Bruccoleri et al. 1983; MacKerell Jr, Brooks III et al. 1998) for all molecular simulation. MF-FEB calculations required the development version of CHARMM due to its dependence on the CHEMPERT module. Additionally, MF-FEB calculations required an “extra-large” build option of CHARMM to provide sufficient memory allocation.

A modified version of Antechamber (Wang, Wang et al. 2006) generated the CHAMM inputs for small organic molecules which were not provided in the CHARMM force field. Dr. Deng (University of Chicago) extended Antechamber’s output to provide the *ligand.crd*, the molecular coordinates, in addition to the *ligand.inp* script files for CHARMM minimization, and *ligand.rtf* and *ligand.prm* files corresponding to the topology and force field parameters for the ligand.

### **2.1.1.3 Molecular Visualization and Manipulation**

We employed three programs to visualize molecular structures: (1) PyMol 0.99 (DeLano 2008) for general visualization, editing targets, and “hand building” small organic molecules; (2) ADT (Morris, Huey et al. 2009) for the preparation and visualization of targets and ligands for docking simulation and reviewing docking simulation results; and (3) VMD ([www.ks.uiuc.edu/Research/vmd/](http://www.ks.uiuc.edu/Research/vmd/)) for visualizing CHARMM .crd files and creating movies of MD trajectories.

We used Openbabel ([openbabel.org](http://openbabel.org)) to convert molecular file formats.

#### **2.1.1.4 Statistics**

Microsoft Excel 2003 and 2007 performed all statistical calculations. Unless otherwise stated, all functions employed the default settings. All t-tests were two-tailed with unequal variance.

#### **2.1.1.5 Scripting**

We wrote and developed all scripts presented in this work using the python programming language versions 2.4 and 2.6 ([www.python.org](http://www.python.org)).<sup>4</sup>

### **2.1.2 Hardware**

Either the Lonestar or the Ranger supercomputer at the Texas Advanced Computing Center or IBM's World Community Grid performed all calculations and simulations presented in this work. The Lonestar high-performance cluster consisted of 1,300 dual-core 64-bit 2.66 GHz Xeon processors (Intel). The Ranger high-performance cluster consisted of 15,744 2.3 GHz AMD Opteron quad-core 64-bit processor. World Community Grid is described in Section 5.2. Work and results were stored at TACC on the "Corral" storage system or at UTMB on a storage system on Random.

### **2.1.3 Molecular File Types**

Multiple file formats exist to store the chemical and structural information of compounds. The .pdb and .mol2 are two common ASCII text based file formats used to store protein and small molecule structures respectively. We choose to use these two file types based on the chemical information they stored, the requirements of the programs used in this work, and personal preference. This subsection provides a brief introduction of both file types. Full

---

<sup>4</sup> All scripts should be run using python version 2.6 to insure proper function. In addition, many of the scripts require the NUMERIC ([sourceforge.net/projects/numpy/](http://sourceforge.net/projects/numpy/)) python package that is not included in the standard python distribution.

description of the .pdb and .mol2 file formats can be found at <http://www.wwpdb.org/docs.html> and <http://tripos.com/data/support/mol2.pdf> respectively.

### ***2.1.3.1 The .pdb File Format***

The .pdb file format was developed to store the experimentally determined coordinates of macromolecular structures in Protein Data Bank (PDB). .pdb files store data based on a fixed format. Each line of the .pdb starts with a tag describing the data stored on that line, and implying the format of that data. Following the tag, each line contains data in a fixed format with each line position containing specified data.

The overall format for a .pdb file starts with the “header” lines containing the title of the structure, details on the experimental method used to determine the structure, the sequence of the structure, the names of any other chemical species resolved in the structure, disulfide bonds for protein structures, and other information. The body of the file contains lines describing the macromolecule’s structure with one atom entry per line.

Three body line types are of particular interest to this work the ATOM, HETATM, and CONNECT lines. ATOM and HETATM share the same format and contain information on each atom in the structure. These lines contain a unique ID number for each atom, the atom name, which residue the atom is a part of, the Euclidian coordinates, the occupancy and b-factor from crystal structures, as well as other information. The difference between the two tags is that ATOM refers to atoms in the macromolecular structure (i.e., the protein), and HETATM refers to everything else (i.e., crystallographic waters/solvent, cofactors, ligands). The CONNECT line contains information on the bonding between HETATM atoms. These lines are placed at the end of the .pdb file. These lines contain no information on bond order and only list bonded atoms by a unique numerical identification corresponding to a HETATM line. Bonds between atoms in the ATOM lines are implied based on amino acid residue and standard atom naming conventions for each residue. (e.g., The alpha, beta carbons in amino acid)

### 2.1.3.2 The .mol2 File Format

The .mol2 file format is a proprietary file format developed by Tripos for its Insight molecular visualization and modeling software. (<http://tripos.com/data/support/mol2.pdf>) It has become a standard format for storing organic small molecule structures. Unlike the .pdb files, .mol2 files presents data in a flexible or “free” format. The .mol2 file is divided into sections beginning with a header and followed by data which is presented using a standard format. The format is defined by the use of white space rather than character position on a line. For example @<TRIPOS>MOLECULE section contains the name of the molecule on one line followed by information on composition and formal charge on a second line separated by white spaces.

@<TRIPOS>ATOM and @<TRIPOS>BOND sections contain the principle structural information. These sections contain lines similar to the ATOM and CONNECT lines in the .pdb file with two differences. First, the lines storing atom position in @<TRIPOS>ATOM contain information on atom type and charge. Second, the lines describing bonding in the @<TRIPOS>BOND section gives the bond order.<sup>5</sup>

## 2.2 PREPARING THE LIGAND AND THE RECEPTOR

The goal of this section is to review the preparation of the ligand and the target files for docking simulation and MF-FEB calculations. Preparation began by obtaining the 3D structures of each molecule and processing the information into the .mol2 file format for the ligands and the .pdb file format for the target. We used these two file formats to generate the input files for both AutoDock4 and CHARMM. While the input files for AutoDock4 and CHARMM are

---

<sup>5</sup> While the white space between data can be a space or a tab, some programs will only correctly read .mol2 files with only one type of white space. It is therefore possible for a correctly formatted .mol2 file to raise an error in program not properly designed. Best practice is to use tabs between data when writing the @<TRIPOS>ATOM and @<TRIPOS>BOND. This practice makes for easier human reading and greater compatibility.

different (requiring different protonation states, atom types, and partial charge assignment) a unified approach to ligand and target preparation provides a common start for both programs and helps avoid errors due to differences in preparation. This section is divided into two parts. Section 2.2.1 covers ligand preparation. Section 2.2.2 covers target preparation.

### **2.2.1 Ligand Prep**

The goal of ligand preparation was to generate a correctly prepared *ligand.mol2* file of the ligand molecule data that was ready for input into both AutoDock4 and CHARMM. A correctly prepared *ligand.mol2* file would contain the 3D coordinates for the ligand, was properly protonated, and each atom and bond was correctly typed and uniquely identified.

The procedure for preparing the *ligand.mol2* is explained in the following two sections. Section 2.2.1.1 discusses the different methods used to obtain coordinates for the ligands. Section 2.2.1.2 outlines methods used to prepare the *ligand.mol2* files for each ligand.

#### ***2.2.1.1 Obtaining Ligand Coordinates***

The first step in ligand preparation was to obtain or generate the 3D atomic coordinates of the ligand. This section examines the three methods used to obtain coordinates, from co-crystal structures, from compound databases, or built by hand.

##### *2.2.1.1.1 Ligand Source: Co-crystal structures*

Co-crystal structures refer to heterogeneous crystal structures such as a ligand bound to a protein, a protein bound to DNA, or a multiple protein complex. For this work, co-crystal ligands refer specifically to small organic molecules bound to a protein obtained from co-crystal structures. We obtained co-crystal structures from the PDB. Atomic coordinates for a ligand were then obtained from the co-crystal .pdb by removing their corresponding HETATM lines using a text editor and generating a new *ligand.pdb* file for the ligand.

Co-crystal ligands were used to test the ability of docking programs to reproduce experimentally verified ligand poses. The main advantage of co-crystal ligands is that their

coordinates have been experimentally determined and are therefore accurate within the resolution of the structure. If the crystallographic resolution is  $> 1 \text{ \AA}$  most hydrogen locations are indeterminable, requiring them to be added by hand as describe below. Using low-resolution co-crystal ligand structures limits RMSD calculations when comparing docking generated poses to experimentally determined poses to only the heavy atoms. Additionally, the starting ligand conformation biases the pose prediction in docking simulations using incremental rotations to model ligand flexibility as it pre-defines the conformational space explored. Starting with a bond conformation therefore overestimates the ability of docking programs to reproduce experimentally determined structures compared to random or minimized starting conformations. Finally the *ligand.pdb* file format does not contain information on bond order in the CONNECT line.

#### 2.2.1.1.2 *Ligand Source: Compound Databases*

As described in Section 1.3.1, different compound databases are available for virtual screening. In this work, we obtained the majority of ligands from the ZINC database (Irwin and Shoichet 2005), including those from the DUD test set (Huang, Shoichet et al. 2006). Ligands obtained from ZINC were downloaded as *ligand.mol2* files, which were ready for use with both AutoDock4 and CHARMM.

Compounds obtained from databases are often correctly protonated, minimized, and contain correct bond order information. However, compound databases have disadvantages. Often, 3D coordinates and other properties are automatically determined due to the large size of the databases. While a vast major of the properties are generated correctly, a small portion of ligands may be incorrect. Also due to the databases' large size, incorrectly prepared ligands are difficult to curate. In general, we found compounds from the ZINC database to be of good quality.

Some compound databases do not enumerate tautomers and stereoisomers for each compound. The ZINC database treated stereoisomers as unique compounds. ZINC stores tautomers as multiple molecular entries in same *ligand.mol2* file. We split tautomer containing files into separate files, one for each tautomeric state, for use with CHARMM and AutoDock4. This separating of files required the addition of a four number tag to the end of all ZINC compound file names to identify the different tautomeric states. Therefore, we identified all ZINC compounds as ZINC#####-####.

#### 2.2.1.1.3 *Ligand Source: Hand-Built*

When compounds were not available as co-crystal ligands or in compound databases, we built the ligand by hand with PyMol's builder function using the standards outlined in Section 2.2.1.2. While hand-building ligands allow any compound to be generated, it is time consuming; therefore, we hand-built only the ligands required for specific calculations. In addition, PyMol could not save files in the propriety .mol2 format; therefore, we saved PyMol generated ligands as a *ligand.mol* file and then converted them to *ligand.mol2* using openbabel.

#### 2.2.1.2 *Preparing Ligand .mol2*

In order to generate input files for AutoDock4 and CHARMM, we prepared each ligand as a *ligand.mol2* file to the following standards. Each ligand was protonated to physiological pH, each atom was uniquely named, and each bond was assigned the correct bond order. This section covers the implementation of each of these standards for coordinates obtained from the sources described in Section 2.2.1.1, with focus on the co-crystal and hand-build sources.

##### 2.2.1.2.1 *Ligand Protonation*

While ligands for the ZINC database were correctly protonated, hand-built ligands and those from low-resolution co-crystal structures required the addition of protons. Unless otherwise noted, we protonated titratable function groups at physiological pH, and if multiple tautomeric forms existed at physiological pH, we generated each form. Protons were added with the automatic protonation functions in ATD or PyMol, and, if needed, edited by hand in PyMol.

#### 2.2.1.2.2 *Ligand Atom Name Assignment*

While not necessary for AutoDock4, CHARMM required that each atom in the ligand be uniquely named; therefore, we built all ligands to this standard. Each atom was assigned a two-part name starting with its elemental symbol followed by a two-digit integer. For example, we labeled the carbon atoms as C01, C02, C03, and so on, while oxygen atoms were labeled O01 and O02. ZINC ligands used this same element-number nomenclature by default. We fixed atom names by editing the *ligand.mol2* file in a text editor or with ligand prep scripts.

#### 2.2.1.2.3 *Ligand Bonds*

ADT depends on the bonding information in the *ligand.mol2* file to correctly assign atom types for their respective force fields and for determining partial charges. While, ZINC *ligand.mol2* contain correct information on bonding, both hand-built and co-crystal ligand *.mol2* files did not automatically have this information. Hand-built ligands gained their bonding information as the molecule was being built in PyMol. We edited co-crystal ligands bonds using PyMol's builder functions to assign correct bond order.

### 2.2.2 **Target Prep**

We obtained the 3D structures of protein targets from the PDB as a *.pdb* file. We prepared the *target.pdb* files using the follow standards to generate inputs for CHARMM and AutoDock4. The *target.pdb* files contained only coordinate lines with only one entry per atom. Missing atoms were added to amino acids with at least backbone atoms present in the determined structure. The ends of each chain were capped as their ions. The protein was protonated to physiological pH and the protons minimized. The following sections detail these steps.

#### 2.2.2.1 *Preparing the .pdb File*

When downloaded from the PDB, a *.pdb* file contained information not required for CHARMM or AutoDock4. We edited the file to contain only "ATOM" and "HETATM" lines with only one entry per atom. Sometimes the downloaded *.pdb* file presented side-chains in multiple conformations due to multiple detectible conformations in the diffraction data. For each

side-chain, we selected only one conformation based on the inspection of non-bonding interactions, particularly H-bonding. If the structure did not suggest a conformation selection, then we selected the conformation with the highest occupancy value. If all poses had the same occupancy, we selected the “A” conformation. Generally selecting the “correct” conformation was not critical because we did not find multiple-conformation amino acids in the docking search volumes, and CHARMM minimized and equilibrated the target during the MF-FEB calculations. None of the targets used in this work contained co-factors. We removed all other molecules from the *target.pdb*, with the exception of the crystallographic waters.

#### ***2.2.2.2 Correcting Missing Atoms***

Amino acids were sometimes missing heavy atoms coordinates. Missing atoms were most common for Lys and Arg residues on the surface of the protein, as their flexibility prevents atomic resolution diffraction. We replaced missing heavy atoms using Pymol’s mutation module by “mutating” the amino acid to itself, thereby replacing the missing atoms. The mutation module allowed for side chain conformation selection. We therefore selected the most sterically favored conformation. For MF-FEB calculations, all missing side-chains were fixed as they provided sources of long-range electrostatic interactions and need to be included in the calculation of the generalized solvent boundary potential. (See 2.4.2.1.2)

#### ***2.2.2.3 Chain Termini***

We capped the ends of all amino acid chains in their ionic form. We protonated the N-terminal amine to the amide. The C-terminus often required modification from an aldehyde to the carboxylic acid by changing the H to an O. We performed the C-terminal modification, and N-terminal protonation using PyMol’s builder module. In structures where sections of the protein were missing, for example NS2B in the dengue protease structures, we did not attempt to rebuild the missing section and ends were left in their ionic form.

#### **2.2.2.4 Protonation**

We added protons to the target and crystallographic waters using ADT, because its proton name conventions corresponded to the naming conventions used by the CHARMM force field for amino acids; however, ADT fully protonated histidine to their cationic state. Therefore, we removed one proton from each histidine using PyMol. The default proton configuration was to leave the delta nitrogen proton, removing the eta nitrogen proton, unless the H-bonding patterns suggested otherwise. After protons were added, CHARMM minimized the protons using adopted basis Newton-Raphson minimization. (Brooks, Bruccoleri et al. 1983; MacKerell Jr, Brooks III et al. 1998)

### **2.3 DOCKING AND VIRTUAL SCREENING WITH AUTODOCK4**

This section describes the methods used for docking and virtual screening with AutoDock4. Section 2.3.1 provides a systematic explanation of how the docking simulations were performed. Section 2.3.2 describes the methods and scripts used to perform virtual screening on TACC's supercomputers, including the preparation of compound databases.

#### **2.3.1 Docking with AutoDock4**

This section describes the steps used to perform a single docking simulation with the AutoDock4 package and ADT.<sup>6</sup> The goal of this section is to outline the methods used for running single docking simulations and provide a context for future method descriptions and discussions.

##### **2.3.1.1 Stage 1: Ligand and Receptor File Generation**

Both AutoGrid4 and AutoDock4 required that the ligand and receptor be in the .pdbqt file format. The .pdbqt file format is an extension of the .pdb file format. In a .pdbqt file the ATOM and HETATM lines are modified by replacing the b-factor and occupancy fields with partial charge (hence “q”) and atom type. The ligand form of the .pdbqt adds torsion (hence “t”)

---

<sup>6</sup> Full tutorials are provided for AutoDock4 and ADT at [autodock.scripps.edu](http://autodock.scripps.edu).

information about the ligand by representing the ligand as a root with a series of moveable branches. The atoms contained in a root or branch section are grouped together in the file with lines before and after the group indicating where they stop and start, nesting groups as needed. Additionally, the header of the ligand .pdbqt file includes a list of all torsions indicating whether they are active, allowed to rotate in the docking simulation, or inactive. We could automatically generate *ligand.pdbqt* using ADT with *ligand.mol2* files prepared to the standards described in Section 2.2.1. However, the *target.pdb* needed their crystallographic waters removed before we could generate *target.pdbqt* files with ADT.

Sections 2.3.1.1.1 and 2.3.1.1.2 discuss the options for and the selections made in ligand and receptor preparation respectively. Section 2.3.1.1.3 is a brief discussion on AutoDock4 receptor flexibility options.

#### 2.3.1.1.1 *Preparing the Ligand for Docking*

ADT automatically processed the *ligand.mol2* into a *ligand.pdbqt* by removing all non-polar protons, assigning AutoDock4 atom types to each of the atoms, assigning Gasteiger partial charges to each atom, and identifying rotatable bonds. AutoDock4 uses a semi-unified atom force field treating only polar protons explicitly; therefore, ADT removed all non-polar protons. The reason we added all protons in the ligand preparation was to maintain conformational consistency in protons between AutoDock4 and CHARMM, which uses an all-atom force field. Next, ADT assigned each atom an AutoDock4 atom type based on element and bonding arrangement.<sup>7</sup> Then, ADT automatically assigned Gasteiger partial charges to a ligand. For hand-built and co-crystal ligands, we used the automatic charging. For compounds from the ZINC database, their native charges, assigned during ZINC preparation, were retained. According to the *ligand.mol2* files from ZINC, ZINC ligands were assigned Gasteiger partial charges, using AMSOL, but upon comparing ADT and ZINC charges, significant difference

---

<sup>7</sup> After automatic processing, aromatic carbons (A) could be modified by hand in ADT.

were noted. After assigning partial charges, ATD automatically detects rotatable bonds and builds a torsion tree for the ligand. AutoDock4 can treat peptide bonds as rotatable bonds. We made all ligand peptide bonds “inactive” and therefore treated them as non-rotatable. All other bonds were assigned rotatable based on the ADT’s automatic preparation. Finally, ATD wrote the ligand as a *ligand.pdbqt* file suitable for docking.

#### 2.3.1.1.2 *Preparing the Target for Docking*

When loaded, ADT automatically processed the *target.pdb* file, with water removed, into in a *target.pdbqt* file. The automatic processing started with ADT removing all non-polar protons from the target. Due to this automation, it was important that the titratable amino acids were correctly protonated before being loaded, as ADT does not automatically add polar protons. After ADT removed all non-polar protons, it added Kollman partial charges to the atoms in the target. Finally, ADT wrote the target as a *target.pdbqt* file, without any torsional information.

#### 2.3.1.1.3 *Using AutoDock4 Receptor Flexibility*

AutoDock4 allowed selected amino acids to be treated flexibly, providing a limited model of target flexibility. Receptor flexibility was implemented by treating selected amino acid side-chains as anchored ligands. The selected side-chains would be added to a *sidechain.pdbqt* file and removed from the *target.pdbqt* file. However upon testing, using the trypsin test set in Section 3.1.1.1, we found that allowing side-chain movement made no improvements in pose prediction, increased search times, and artificially lowed the docking score. Therefore, we did not employ this feature of AutoDock4.

#### 2.3.1.2 *Stage 2: Scoring Grid Generation*

AutoDock4 uses scoring grids to decrease the computational expenses of scoring and to define the search volume. AutoGrid4 built a scoring grid for each atom type in a given ligand by incrementally moving each atom type through the search volume and scoring each point in the cubic volume. The default grid spacing was  $\sim 1/3$  Å. We defined the search area using ADT, which allows for visualization of the search volume overlaid on the target. We determined the

search volume by centering it on a co-crystal ligand or a binding site. Then, we expanded or compressed the volume of the box to encompass the whole ligand and at least 2-3 Å beyond the ligand to insure that all target ligand interactions were included in the search volume. For virtual screening, we extended the search volume to include key features of a binding site. For example, with trypsin, we included the P1-P4 , P1' pockets and the active site in the search volume. These large search volumes were preferred for virtual screening because the intention of our viral screening projects was to find any compound that would bind in a given region of the target and not to reproduce known target-ligand interactions. Once we defined the area, ADT generated the grid parameter file (*job.gpf*) which contained instructions for AutoGrid4 on how to construct the scoring grids.

We ran AutoGrid4 through ADT or on the command line. AutoGrid4 required the *target.pdbqt*, the *job.gpf* file, and a parameter file for the AutoDock4 force field (*AD4\_parameters.dat*) to run. The command to run AutoGrid4 was “autogrid4 -p *job.gpf* -l *job.glg*”. AutoGrid4 generated one file for each atom type in the ligand, and a *target.xyz* file, all of which were required to run AutoDock4. The *job.glg* was a log file that contained the output from AutoGrid4, and was used to check for and troubleshoot errors.

### **2.3.1.3 Stage 3: Docking Simulation**

AutoDock4 performed the docking simulation, which we ran through ADT or on the command line. AutoDock4 uses all the output files generated from AutoGrid4 except the *job.glg* file. It also requires the *ligand.pdbqt*, *target.pdbqt*, and a parameter file for the AutoDock4 force field (*AD4\_parameters.dat*). Like AutoGrid4, AutoDock4 uses a parameter file for runtime instructions (*job.dpf*). We generated the *job.dpf* using ATD. The *job.dpf* included information on the ligand, target, the grid files, and the search parameters. Docking simulations used the Lamarckian genetic search algorithm with a SW local optimization. (See Section 1.2.4) The default values were used for all parameters, except, we used a population size of 200 members, a maximum number of energy evaluations (MNEE) of 1 million, and 100 GA runs. The command

to run AutoDock4 was “autodock4 -p *job.dpf* -l *job.dlg*” with all the docking results written to the *job.dlg*.

#### **2.3.1.4 Stage 4: Results Analysis and Visualization**

AutoDock4 stored the docking results, the poses, and scores in the *job.dlg* file, which ATD could visualize. Additionally, AutoDock4 atomically performed cluster analysis on the docking results grouping similar poses, using an RMSD cutoff of 1 or 1.5 Å. The *job.dlg* file contained the results of the cluster analysis.

### **2.3.2 Virtual Screening with AutoDock4**

This section describes the methods and scripts used to perform virtual screening using AutoDock4 at TACC. We used these methods to run virtual screening on the five DUD targets described in Chapter 3. This section is divided into three sections. Section 2.3.2.1 describes the preparation of the compound database for virtual screening. Section 2.3.2.2 describes how the virtual screenings were performed. Section 2.3.2.3 describes how the results were extracted.

#### **2.3.2.1 Compound Databases – Source and Generation**

In order to perform virtual screening using AutoDock4 we built compound databases of *ligand.pdbqt* files to the standards described in Section 2.3.1.1.1. The source of these compounds was the ZINC database. The files downloaded from ZINC came as multi-compound *.mol2* files. We divided multi-compound *.mol2* files into individual compound *ligand.mol2* files and then processed the *ligand.mol2* files into *ligand.pdbqt* files. The final product of the processing was two directories, one containing all the *ligand.mol2* files and a second containing all the *ligand.pdbqt* files. We used the *ligand.pdbqt* directory for virtual screening, but retained both directories for MF-FEB calculations as described in Chapter 6.

The *zincextractor3.py* script took a single GNUzipped multi-compound *.mol2* file from ZINC and separated it into individual *ligand.mol2* files. The script processed each file separately, but individual scripts were executed as a single batch job on Lonestar or Ranger to

process multiple multi-compound .mol2 files. The script worked by reading through the GNUzipped multi-compound .mol2 file finding the start of each .mol2 file based on the standard .mol2 molecule tag and extracting the ligand. The script wrote each extracted *ligand.mol2* file to the output directory using the extended ZINC ID as described in Section 2.2.1.1.2. Since ZINC does not uniquely name the tautomeric states, the script kept a list of all ligands extracted and would incrementally number the last four characters (0000, 0001, 0002 ...) of the file name each time a ligands name in the GNUzipped multi-compound .mol2 file appeared, thereby writing each tautomer as a separate file.

The *mol2topdbqt\_v01.py* script read each *ligand.mol2* file from one directory and processed the file to write out a *ligand.pdbqt* file, ready for docking, into a new directory. The script processed each compound by calling the same python modules used by ADT for ligand preparation. These modules were the PyBabel, PyAutoDock, MolKit, and AutoDockTools modules from ATD, although only MolKit's Read and AutoDockTools.MoleculePreparation's AD4Ligand Preparation function were called directly from *mol2topdbqt\_v01.py*. The script processed the ligands as described in Section 2.3.1.1.1 .

### **2.3.2.2 Virtual Screening at TACC**

To perform a virtual screening, the *cf-100128.py* script processed each compound in the *ligand.pdbqt* library into one complete docking job that was then run on Lonestar. The *cf-100128.py* script required two input directories. The first was the compound database, named "cdb". The compounds in the compound database were grouped into alphabetic subdirectories (i.e., a,b,c, etc.) according to the size of the compound database. The "a" group, for example, also contained a number of subdirectories corresponding to docking "jobs" which contained collections of ligands that were to be run at the same time. The second directory, "wgsrc", had a sub directory, named for the target system, containing the *target.pdbqt*, template versions of the *job.gpf* and *job.dpf* files, and the *AD4\_parameters.dat* file. The *target.pdbqt* was prepared as described in Section 2.3.1.1.2. We made the template versions of the *job.gpf* and *job.dpf* by first

generating the files using ADT, consistent with the virtual screening parameters described in Section 3.1.1.3, for a ligand. We then removed the ligand specific information from the *job.gpf* and *job.dpf* files generating the template versions.

For each *ligand.pdbqt* file in the compound database input directory, *cf-100128.py* generated a new directory containing all required files and a shell script to run AutoGrid4 and AutoDock4. To prepare docking runs, the *cf-100128.py* script first read in the target and parameter files generating an object. The script gave the ligand object to generate the *job.gpf* and *job.dpf* files for each ligand. The object then built a shell script to start all AutoGrid4 and AutoDock4 runs for each ligand and copied all required files into newly created job directory. The *cf-100128.py* script used the same python scripts the ADT used for generating *job.gpf* and *job.dpf* files, the PyBabel, PyAutoDock, MolKit, and AutoDockTools modules. The script packaged all the docking job directories into a single *.tar.gz* file and wrote the job scripts for submitting a job on Lonestar.

We performed virtual screening on Lonestar using the *LSF\_launcher*, which runs multiple single processor jobs on a set of nodes. Both Lonestar and Ranger have programs to run multiple single processor jobs on a set of nodes. In both cases, the launcher programs required one file listing all the jobs to be run and a second script file with the job submission command.

When the screening was complete, we copied the *job.dlg* files into a single directory for analysis.

### **2.3.2.3 Results Processing**

The *dlg\_reader\_05.py* script processed virtual screening results. The script read each *job.dlg* file and selected the best pose using the standard pose selection method described in Section 3.1.1.2. As each *job.dlg* file was read in, the script built an object, using the *docking\_run.py* module, containing the docking results. The *docking\_run.py* module was based on ADT results processing tools. The module performed the cluster analysis anew when a

*job.dlg* was fully read in. The script then wrote the best poses for each compound in the virtual screening into a tab delineated table text file with one compound record per line. Each line contained the compound ID, cluster size, score breakdown, and the pose. The pose was stored as the coordinates for the geometric center of the compound, the x, y, z translation for the compound with reference to the target, the axis angle rotation data for the compound, and the torsion angle for each rotatable bond. The pose was not stored as atomic coordinates to save space and to facilitate MF-FEB calculations.

## **2.4 MF-FEB CALCULATIONS**

This section describes the methods used to perform MF-FEB calculations using TACC's high performance computers Lonestar and Ranger. Section 2.4.1 describes the preparation of the ligand and target for MF-FEB calculations. Section 2.4.2 describes how to perform MF-FEB calculations using CHARMM. Section 2.4.3 describes how MF-FEB calculations were run on Lonestar and Ranger.

### **2.4.1 Target and Ligand Preparation**

In order to perform MF-FEB, we prepared the ligand and the target files to be used with CHARMM. Section 2.4.1.1 describes ligand file preparation. Section 2.4.1.2 describes target preparation. Section 2.4.1.3 briefly describes the water model used in the simulations.

#### ***2.4.1.1 Ligand Preparation***

To prepare the ligand files for MF-FEB calculations, we first generated *ligand.mol2* files according to the standards outlined in Section 2.2.1.2, whose coordinates corresponded with the desired ligand pose, whether determined experimentally or generated from docking. Antechamber then processed the *ligand.mol2* into CHARMM input files. This section outlines the methods used to generate the correctly posed *ligand.mol2* files and generate the CHARMM input files.

We either generated correctly posed *ligand.mol2* files automatically or built it by hand. The procedure for automatic generation is discussed in Chapter 6. We used the methods described in Sections 2.2.1.1.1 and 2.2.1.2 for hand building the *ligand.mol2* files for ligands with co-crystal structure poses. To build by hand the *ligand.mol2* from docking results, we removed the .pdb style coordinate lines for the desired pose from the *job.dlg* file using a text editor and saving the extracted lines to a *ligand.pdb* file. We processed the *ligand.pdb* file in the same manner as a co-crystal pose. In both cases, CHARMM minimized the ligand's protons, but not the heavy atoms. Additionally, we transposed the coordinates of the ligand so that the geometric center of the ligand was at the origin. The adjustments used for the transposition were retained to transpose the target coordinates.

The CHARMM force field was designed to simulate macromolecules and does not contain the force field parameters for the ligand molecules. Therefore, to simulate the ligand in CHARMM a *ligand.mol2* needed to be parameterized and converted into the three input files required for CHARMM. These input files were *ligand.crd* containing the atomic coordinates, a *ligand.prm* containing the force field parameters for the ligand, and a *ligand.rtf* containing the topological information for the ligand in the force field. We prepared the CHARMM input files using the modified version of Antechamber, parameterizing the ligand with the CHARMM compatible generalized AMBER force field (Wang, Wolf et al. 2004), and to assign partial charges using the AM1-BCC charge method (Jakalian, Jack et al. 2002).

#### ***2.4.1.2 Target Preparation***

The goal of receptor preparation is to generate the files need to simulate the receptor in CHARMM. The files required include the receptor coordinates (*target\_xxxx\_water.crd*) and two CHARMM stream files that contain the amino acid sequence (*squ.str*), and di-sulfide bonds and chain caps information for adjusting the non-standard amino acids (*disu.str*). This section presents the method and scripts used for generating the target file starting with *target.pdb* generated using the standards described in Section 2.2.2.

Before generating the CHARMM target inputs, we performed preparatory steps. First, because the CHARMM force field treats each histidine tautomer as a unique amino acid, we determined and recorded the protonation states of each histidine residue. Next, we returned the crystallographic waters to the *target.pdb*, if they had been removed for docking. Then, we generated the *disu.str* in a text editor. The *disu.str* contained the information for CHARMM to build the di-sulfide bonds and to modify chain ending amino acids. Finally, we fixed the O in the C-terminal carboxylic acids and the N-terminal amine H names to correspond to naming convention used by the CHARMM force field.

To prepare the *target\_xxxx\_water.crd* and *seq.str* files, we developed the *pdertools.py* module. The module was called by other scripts to prepare target *.pdb* and *.crd* files as needed. The module first read the *target.pdb* file generating an object containing all the atoms in the target. Then the object was passed to other module functions for modification. First, the atom and residues were renumbered starting at one. Then, the histidines were modified using a directory (i.e., hash table) of the residue number and histidine type prepared from the data collect during preparation. Next, the coordinates were transposed according to the information obtained in ligand preparation. Then, the protein sequence was extracted and written as the *seq.str* file. Finally, the *rept\_xxxx\_water.crd* was written. The module had a function to prepare files for minimization of the target using CHARMM. The minimization outputs were used to test the CHARMM target input files for errors.

### **2.4.1.3 Solvent**

We modeled all explicated waters as TIP3 waters as provided in the CHARMM force field files. (Mahoney and Jorgensen 2000)

### **2.4.2 MF-FEB Calculations**

As described in Section 1.4.4, MF-FEB can calculate the FEB using molecular mechanics methods. We employed two “generations” of CHARMM scripts to perform MF-FEB

simulations and calculations. Drs. Deng and Roux of the University of Chicago kindly provide the original CHARMM scripts used in their MF-FEB calculations.(Deng and Roux 2006; Deng and Roux 2008) These original scripts are referred to as the first generation scripts. We developed a second generation of scripts for performing MF-FEB calculations on World Community Grid. The second generation simplified and refined the first generation CHARMM scripts and added check-pointing functions required for World Community Grid.

This section describes performing MF-FEB calculations in CHARMM using the second generation of scripts and is divided into three subsections. Section 2.4.2.1 describes the “site” portion of the MF-FEB calculations. Section 2.4.2.2 describes the “bulk” or “solvent” portion of the MF-FEB calculations. Finally, Section 2.4.2.3 describes the calculation of the FEB from the perturbation simulations.<sup>8</sup>

#### ***2.4.2.1 Site***

The goal of the site simulations was to model the decoupling of the ligand from the site under positional and conformational constraints, and the subsequent release of those constraints. This section describes the methods used for site simulations in MF-FEB calculations. Section 2.4.2.1.1 describes the site system. Section 2.4.2.1.2 describes the steps for preparing the system for perturbation simulations. Section 2.4.2.1.3 concludes with a description of the perturbation simulations.

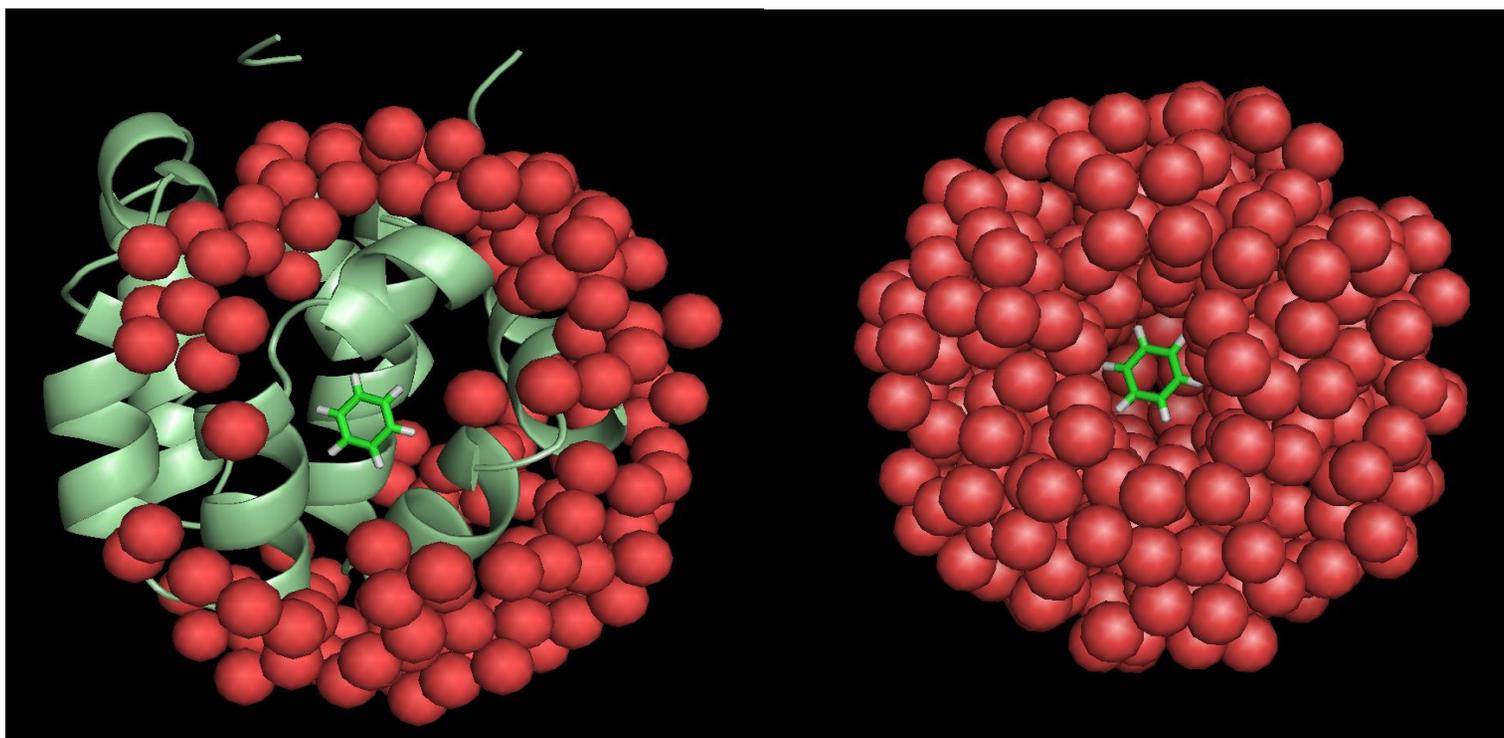
##### ***2.4.2.1.1 The System***

The site system modeled the ligand in the target’s binding site. Figure 2-1 presents an image of the system. We modeled the site as a solvent drop with a 18 Å radius centered on the ligand with all atoms within the radius treated explicitly and all atoms beyond the sphere represented using a generalized solvent boundary potential (GSBP).

---

<sup>8</sup> Parameters in these sections also apply to MF-FEB calculations on World Community Grid.

Figure 2-1 – The “site” (left) and “bulk” (right) systems used for MF-FEB calculations for the binding of benzene to the L99A T4 lysozyme engineered hydrophobic binding site. The red spheres represent the locations of water molecules. Some waters were removed to show ligand and target. The mint green ribbon represents the backbone of the target. The stick model in the center represents benzene with green for carbon atoms and white for hydrogen atoms.



#### 2.4.2.1.2 System Preparation, Grand Canonical Monte Carlo and System Equilibration

To prepare a site for perturbation simulations required three steps. First, we constructed the GSBP. Second, we equilibrated the number of waters in the site using a grand canonical Monte Carlo (GCMC) simulation. Third, we equilibrated the site's atoms using MD and selected the positional constraints. This section describes each of those steps.

CHARMM used the instructions in the `site_stup_gsbp.inp` file to prepare the system and the GSBP. The script required the following input files: `target_xxxx_water.crd`, `seq.str`, and `disu.str` files from target preparation; the `ligand.crd`, `ligand.rtf`, and `ligand.prm` files from ligand preparation; the `par_all22_prot_cmap.inp` and `top_all22_prot_cmap.inp` files containing the CHARMM force field; and the `water_8000.crd_nocopy` file containing an equilibrated 64 Å x 64 Å x 64 Å water box.<sup>9</sup> The first step in `site_stup_gsbp.inp` script was to calculate the long-range corrections for the VdW interaction. To begin, CHARMM built the site system using the `.crd` and force field parameter files. It then added the water box and then removed all waters overlapping the target, the ligand, and the crystallographic waters. Next, CHARMM determined the potential energy of the VdW interaction of the protein and ligand using a non-bond interaction cutoff of 1000 Å to include all interaction in the system. Then, CHARMM removed all solvent atoms beyond 3 Å and all protein atoms beyond 7 Å from the GSBP radius. CHARMM calculated the VdW potential energy between the target and ligand atoms using a non-bonding cutoff equal to the GSBP sphere radius. Finally, the difference in the two potential energies was the long-range correction value. The second step calculated GSBP. CHARMM started by rebuilding the system as described above using the water box and deleting all overlapping molecules and all waters beyond the GSBP. Next, it set the radii of the inner and outer shells, where the inner shell is where all atoms are explicit, and the outer shell is 7 Å

---

<sup>9</sup> The `const_dihe.str` file was included in this and all other steps for use in modeling symmetric ligands, but we did not utilize this function, therefore the file is not listed in later sections.

beyond where only target atoms are retained. Then, CHARMM generated two fields using the GSBP function in the CHARMM Poisson-Boltzmann equation solver module. The first field was the generalized reaction field, modeling the region between the inner and outer shell radii. The second field was the static external field, modeling everything beyond the outer shell radius. At the end of the GSBP set-up process CHARMM had generated several files. The first file was the *fe\_lrc\_site* file<sup>10</sup> that contained the long-range correction used in determining the FEB. Next were the files defining the GSBP. These files include the *target\_ligand\_gsbp\_param\_rad.str* file containing the GSBP parameters; and the *target\_mij.dat* and *target\_phi.dat* binary files contain the reaction and static interaction grids respectively. The final files, *target\_ligand\_gsbp\_rad\_0.crd* and *target\_ligand\_gsbp\_rad\_0.psf*, contained the coordinates and system information to quickly rebuild the system.

CHARMM equilibrated the number of waters in the site by GCMC using the *site\_stup\_gcmc.inp* script. This stage required the following input files: *target\_ligand\_gsbp\_rad\_0.crd* and *target\_ligand\_gsbp\_rad\_0.psf* files containing the system coordinates and information; the *target\_ligand\_gsbp\_param\_rad.str*, *target\_mij.dat*, and *target\_phi.dat* to set-up the GSBP; and the *par\_all22\_prot\_cmap.inp*, *top\_all22\_prot\_cmap.inp*, *ligand.prm*, and *ligand.rtf* files for the force field. The GCMC equilibration began by rebuilding the system with the GSBP. Then, CHARMM minimized the system using CHARMM's steepest decent minimization method for 1000 steps. Next, CHARMM added ghost waters forming a reserve of waters for the GCMC. Following reservoir construction, CHARMM carried out the GCMC simulations as previous described. (Deng and Roux 2008) In short, the process began with the GCMC simulations where the waters in the system were equilibrated, where if space was sufficient to accept a new water, a ghost water would be add, or if the packing was too tight, a water would be removed, hence grand canonical. A  $\mu$  value of  $-4.25 \text{ kcal mol}^{-1}$  determined the acceptance or rejection of a water addition or removal. After 10,000 MC steps, CHARMM

---

<sup>10</sup> This is a txt file with no extension.

equilibrated the system using Langevin MD for 10 ps at 300K with 2 fs time steps and the non-bond cutoff equal to the radius of the GSBP sphere. CHARMM repeated the MC/MD cycle 10 (or 50) times to equilibrate the waters. The output files *target\_ligand\_gsbp\_rad\_1.crd* and *target\_ligand\_gsbp\_rad\_1.psf* contained the water-equilibrated system.

The final stage of site preparation was MD equilibration of the system followed by ligand constant selection and RMSD fluctuation calculation. The *site\_stup\_eqlb.inp* and the *pick.str* scripts contained the CHARMM instructions for this stage, which required the following input files: *target\_ligand\_gsbp\_rad\_1.crd* and *target\_ligand\_gsbp\_rad\_1.psf* for the system; *target\_ligand\_gsbp\_param\_rad.str*, *target\_mij.dat* and *target\_phi.dat* to set-up the GSBP; and the *par\_all22\_prot\_cmap.inp*, *top\_all22\_prot\_cmap.inp*, *ligand.prm*, and *ligand.rtf* files to define the force field. After the system was constructed, CHARMM minimized the system with 1,000 step steepest decent minimization followed by 1,000 step adopted basis Newton-Raphson minimization. (Brooks, Bruccoleri et al. 1983; MacKerell Jr, Brooks III et al. 1998) Then, Langevin MD simulations ran for 200 ps at 300 K with a 2 fs time step to equilibrate the system with a non-bond cutoff distance equal to the GSBP sphere radius. Next, CHARMM selected the ligand's positional restraints by randomly selecting atoms in the ligand and target to serve as points of constraint. CHARMM accepted or rejected the randomly selected constraints based on the relative geometry between the three selected points in the ligand and the three selected points in the target. Finally, CHARMM determined the average position for the center of mass of the ligand and its RMSD fluctuation for the last 180 ps of MD trajectories. CHARMM outputted the constraint definitions as the *const\_para\_target\_ligand\_equ.str* and *restr\_def.str* stream files. Additionally, the output *target\_ligand\_gsbp\_rad\_2.crd* file constrained the final pose of the system after MD simulations and *target\_ligand\_ave.crd* contained the average system coordinates from the MD run. The output *target\_ligand\_ave\_lig.crd* and *target\_ligand\_eq\_lig.crd* files contained the average and equilibrium coordinates for the ligand

respectively. In addition, the *target\_ligand\_rmsd\_fluc.rmsd* and *target\_ligand\_para\_fluc.rmsd* contained the ligand RMSD fluctuation data.

#### 2.4.2.1.3 Perturbation Simulations

The goal of the perturbation stage was to obtain the microstate potential energies from MD simulations as CHARMM decoupled the ligand from the site. For site simulations, the decoupled properties were electrostatic interactions, VdW dispersion, VdW repulsion, positional constraints, and conformational constraint. The lambda steps for decoupling the electrostatic interaction were: 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1. The lambda steps for decoupling for the VdW dispersion interaction were: 0.0, 0.25, 0.5, 0.75 and 1. The lambda steps for decoupling the VdW repulsion interaction were 0 and 1, but soft-core potentials were used which were varied in steps of: 0.0 to 0.2, 0.2 to 0.3, 0.3 to 0.4, 0.4 to 0.5, 0.5 to 0.6, 0.6 to 0.7, 0.7 to 0.8, 0.8 to 0.9, and 0.9 to 1.0. The lambda steps for removing the positional constraints were: 0.0, 0.0025, 0.005, 0.0075, 0.01, 0.02, 0.04, 0.06, 0.08, 0.1, 0.2, 0.4, 0.6, 0.8, and 1. The steps for removing the conformational constraint on the ligand corresponded to the allowed RMSD fluctuation of the ligand during the MD simulations. The conformational constraint stages steps were: 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, and 5.0 Å. For each step of each stage, we performed two MD simulations running the perturbation step first in one direction and then another. Therefore, we ran each perturbation stage from a lambda value of 0 to 1 and then from 1 to 0.

The goal of each perturbation step was to sample the energy of the microstates generated by a MD simulation for a single lambda step, or RMSD constraint. The *site\_pert.inp* script contained the CHARMM's instructions for this phase. The *site\_pert.inp* script required the following input files: the *target\_ligand\_gsbp\_rad\_1.psf* and *target\_ligand\_gsbp\_rad\_2.crd* files for the system; the *target\_ligand\_gsbp\_param\_rad.str*, *target\_mij.dat*, and *target\_phi.dat* files to set-up the GSBP; the *par\_all22\_prot\_cmap.inp*, *top\_all22\_prot\_cmap.inp*, *ligand.prm*, and *ligand.rtf* files for the force field; the *restr\_def.str*, *const.str*, *target\_ligand\_ave\_lig.crd*, and

rstr\_def\_rms\_fluc.str files for the constraints definitions; and for all perturbation steps except conformation constraints the *target\_stage\_lambda.prt* file containing the perturbation definitions. The perturbation simulations started by building the system and applying the positional and conformational constraints to the ligand. After the system was constructed, CHARMM minimized the system with 1,000 step steepest decent minimization followed by a 1,000 step adopted basis Newton-Raphson minimization (Brooks, Bruccoleri et al. 1983; MacKerell Jr, Brooks III et al. 1998). Next, CHARMM's PERT module modified the system according to the stage and lambda step (i.e., electrostatic interaction at a lambda of 0.1). We utilized the CHEM command from the PERT module. The CHEM command removes the need for additional vacuum simulations to calculate inter-molecular interaction in the ligand by maintaining the interactions during decoupling. In the case of releasing conformation constraints, CHARMM modified the conformational constraints on the ligand according to the allowed RMSD fluctuation. Following perturbation, Langevin MD simulations equilibrated and then sampled the microstates. The Langevin MD ran for different lengths depending on the stage. However, all MD simulation ran at 300 K, with a 2 fs time step, with a non-bond cutoff distance equal to the GSBP sphere radius, and with potential energies extracted every picosecond for analysis. The sampling times for site dispersion, repulsion, and electrostatic stages were 120 ps with 40 ps equilibration. The sample times for positional and conformation constraint stages were 60 ps with 20 ps and 40 ps equilibration stages, respectively. The only file retained from the perturbation stage was the *target\_ligand\_stage\_lambda.wham* file, or the *target\_rmsd\_lambda.rms* file in the case of the conformation constraints, containing the sampled potential energies.

#### **2.4.2.2 Bulk**

The goal of the bulk, or solvent, simulations was to model the coupling of the ligand into the bulk solvent. This section describes the methods used for the bulk simulations in MF-FEB calculations. Section 2.4.2.2.1 describes the bulk system. Section 2.4.2.2.2 describes the

equilibration of the bulk system. Section 2.4.2.2.3 concludes with a description of the perturbation simulations.

#### 2.4.2.2.1 *The System*

The bulk system modeled the ligand free in solvent. Figure 2-1 presents an image of the system. We modeled the bulk as a solvent drop with an 18 Å radius centered on the ligand with all atoms within the radius treated explicitly and the solvent beyond the radius with a spherical solvent boundary potential.

#### 2.4.2.2.2 *System Equilibration*

CHARMM equilibrated the bulk system using the `solv_stup_eqlb.inp` script. The `solv_stup_eqlb.inp` script required the following input files: the `ligand.crd` and `water_400.crd` files containing the atomic coordinate for the ligand and bulk solvent respectively; and the `par_all22_prot_cmap.inp`, `top_all22_prot_cmap.inp`, `ligand.prm`, and `ligand.rtf` files for the force field. CHARMM built the bulk system by placing the ligand at the center of the pre-equilibrated water box defined by the `water_400.crd` file, removing waters that overlap the ligand or were outside the radius of the solvent drop, and adding the SSPB. After the system was constructed, CHARMM minimized the system with 1,000 step steepest decent minimization followed by a 1,000 step adopted basis Newton-Raphson minimization. (Brooks, Bruccoleri et al. 1983; MacKerell Jr, Brooks III et al. 1998) Then, Langevin MD simulations ran for 40 ps at 300 K with a 2 fs time step to equilibrate the system with a non-bond cut off equal to the SSPB radius. CHARMM saved the equilibrated bulk system as `solv_ligand.psf` and `solv_ligand_eq.crd` files.

#### 2.4.2.2.3 *Perturbations*

The goal of the perturbation stage was to obtain the microstate potential energies for MD simulations as CHARMM decoupled the free ligand from the bulk solvent. For bulk simulations, the decoupled properties were electrostatic, VdW dispersion, and VdW repulsion interactions. The lambda steps for decoupling the electrostatic interaction were: 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1. The lambda steps for decoupling the VdW dispersion interaction were: 0.0,

0.25, 0.5, 0.75 and 1. The lambda steps for decoupling the VdW repulsion interaction were 0 and 1, but soft-core potentials were used which were very in steps of: 0.0 to 0.2, 0.2 to 0.3, 0.3 to 0.4, 0.4 to 0.5, 0.5 to 0.6, 0.6 to 0.7, 0.7 to 0.8, 0.8 to 0.9, 0.9 to 1.0. For each step of each stage, we performed two MD simulations running perturbation step in one direction and then another. Therefore, we ran each perturbation stage from a lambda value of 0 to 1 and then from 1 to 0.

The goal of each perturbation step was to sample energy of the microstates generated by a MD simulation for a single lambda step. The *solv\_pert.inp* script contained the CHARMM's instructions for this phase. The *solv\_pert.inp* script required the following input files: the *solv\_ligand.psf* and *solv\_ligand\_eq.crd* files to define the system; the *par\_all22\_prot\_cmap.inp*, *top\_all22\_prot\_cmap.inp*, *ligand.prm*, and *ligand.rtf* files for the force field; (the *const\_dihe.str* and *const\_para\_ligand\_target\_equ.str* for symmetric ligands); and the *solv\_stage\_lambda.prt* file containing the perturbation definitions. After the system was constructed, CHARMM minimized the system with 1,000 step steepest decent minimization followed by a 1,000 step adopted basis Newton-Raphson minimization. (Brooks, Bruccoleri et al. 1983; MacKerell Jr, Brooks III et al. 1998). Next, CHARMM's PERT module modified the system according to the stage and lambda step. Following perturbation, Langevin MD simulations equilibrated, and then sampled the microstates. The MD equilibrated the system for 40 ps and then sampled for 80 ps at 300 K, with a 2 fs time step, with a non-bond cut off equal to the SSBP radius, and with potential energies extracted every picosecond for analysis. The only file retained from the pert stage was the *solv\_stage\_lambda.wham* file containing the sampled potential energies.

### **2.4.2.3 Results**

We obtained the FEB from the MF-FEB simulations in four steps. First, we used weighted histogram analysis method (WHAM) (Kumar, Rosenberg et al. 1992) to calculate the energy contribution to the FEB of the electrostatic, VdW dispersion, VdW repulsion, decoupling of the ligand from the site and bulk, and releasing the ligand's positional constraints. CHARMM analyzed the compiled .wham file outputs from each perturbation stage, checked for convergence

and calculated the energy. Second, we determined  $F_t$  and  $F_r$  numerically. We calculated energy from  $F_t$  by integrating, using a Simpson approximation, the change in force as defined in equation 1-22 as distance using the two angles defining the translation constants defined by the six selected constant atom and extended the distance to 100 Å and the angles rotated 180°. We calculated energy from  $F_r$  using the same method as  $F_t$  only using the three angles that defined the rotational constants. Third, we determined the energy contribution from conformational restraints by solving equation 1-23 using WHAM analyzed data from the MD data obtained from the conformation constraint stages. Forth, we compiled all the energies obtained in the first three steps and the long-range VdW correction. The FEB was the difference between the sum of energies from decoupling the charge-charge, VdW repulsion, and VdW dispersion interaction of the ligand and the site, the energy for releasing the ligand from the constants composed of the energy from the constraint simulations and  $F_t$  and  $F_r$ , and the long range VdW correction; and the sum of decoupling, the charge-charge, VdW repulsion, and VdW dispersion interaction of the ligand and the bulk.

### **2.4.3 Scripts for MF-FEB Calculations on TACC**

This section describes the process of running MF-FEB calculations on the Ranger or Lonestar supercomputers. We developed this process to utilize the first generation of CHARMM scripts, therefore the exact input and output files vary from those described in Section 2.4.2. However, the MF-FEB calculations are identical to those described in Section 2.4.2. Section 2.4.3.1 describes using the `dg_prep_launcher2_01.py` script to build MF-FEB jobs for submission on Ranger or Lonestar. Section 2.4.3.2 describes running the jobs and processing the results.

#### ***2.4.3.1 Building the Jobs***

We developed the `dg_prep_launcher2_01.py` script to prepare multiple MF-FEB calculations to run on Ranger or Lonestar. The script processed three input directories containing the files required for the MF-FEB calculations and generated a new directory containing the MF-

FEB jobs. This section describes the preparation of the three source directories and the structure of the script's output directory.

The `dg_prep_launcher2_01.py` script required three input directories. The first contained all files for running CHARMM. For reference, the files were `const.str`, `par_all22_prot_cmap.inp`, `pick.str`, `radius.str`, `restr_sel.str`, `rstr_def_rms_fluc.str`, `sitp_pert.inp`, `site_stup_eqlb.inp`, `site_stup_gcmc.inp`, `site_stup_gspb.inp`, `solv_pert.inp`, `solv_stup_eqlb.inp`, `top_all22_prot_cmap.inp`, `water_400.crd`, `water_8000.crd_nocopy`, and `wham.inp`. A second directory contained `fe.pl`, used for calculating the results. The third directory, the “jobs” directory, contained subdirectories for each MF-FEB calculation. Each subdirectory contained three items. The “rep” directory containing the *target\_ligand\_water.crd*, `seq.str`, `disu.str`, and `const_dihe.str` generated during receptor preparation as described in Section 2.4.1.2. The “lig” directory containing the *ligand.crd*, *ligand.rtf*, and *ligand.prm* generated during ligand preparation as described in Section 2.4.1.1. The final item was *job.xml* file, whose name matched the MF-FEB calculation's subdirectory name. The *job.xml* file contains all the parameters for the MF-FEB calculations.

The `dg_prep_launcher2_01.py` script built the MF-FEB jobs that corresponded to the jobs subdirectories, by building dictionaries, copying files, and writing out job submission and shell scripts. The script required the `xmltodir.py` modules, which read in the *job.xml* file and processed it into a python dictionary object. The `dg_prep_launcher2_01.py` script generated a single directory containing all the MF-FEB calculations and the scripts to run them. The first level of the output directory contained one subdirectory for each MF-FEB calculation. The directory also contained the scripts for submitting jobs and collecting the results. Each subdirectory contained directories for site and bulk set-up; a directory for each perturbation stage for both systems; three source directories containing the CHARMM inputs, ligand information, and the GSBP file; and the scripts for running the MF-FEB calculations and determining FEB. The script created links between the source directories and the set-up and perturbation directories

to avoid multiple copies of some files. The perturbation and set-up directories contained all the files required to run CHARMM for all perturbation and set-up steps. The directory also contained shell scripts to run each CHARMM simulation using the parameters in the *job.xml* file.

#### ***2.4.3.2 Running and Processing the Results Collection***

As described in Section 2.3.2.2, Ranger and Lonestar had “launcher” programs that allowed the submission of multiple serial jobs to a set of nodes. We designed the *dg\_prep\_launcher2\_01.py* script to build job submission scripts and to create jobs that efficiently balanced the use of Ranger and Lonestar’s resources. Running MF-FEB calculations took four steps.

First, we ran the set-up stages for the site and bulk simulations. We used two run scripts to submit all the jobs, one for the site systems and a second for the bulk systems. The jobs were shell scripts that contain the commands to run CHARMM. For example, the site set-up shell script would run the CHARMM scripts to generate the GSBP, run the GCMC water equilibration, and run the MD system equilibration. In the case of 50 GCMC cycles, we divided the set-up shell script into two sequential parts each with 25 GCMC cycles due to the 24-hour run time limit on Ranger. The *dg\_prep\_launcher2\_01.py* script would build two site set-up run scripts of the spilt site set-up.

The second and third stage was the perturbation simulation and WHAM calculations. We handled job summations for both steps in the same way. Each MF-FEB calculation had job scripts for site and bulk perturbation simulations and WHAM calculations, making four scripts in total. Each of the job scripts ran shell scripts with the CHARMM commands for each perturbation step in the MF-FEB calculations. The first level of the *dg\_prep\_launcher2\_01.py* script’s output directory contained shell scripts that would submit all the job scripts to Ranger or Lonestar.

Finally, we generated the FEB from the WHAM results using the fe.pl script. Dr. Deng (University of Chicago) developed the fe.pl to perform the same function of the dg\_prep\_launcher2\_01.py script previously described. The fe.pl script also contained the result calculating functions, therefore we used it to obtain the FEB from the MF-FEB calculations. We used shell scripts generated by the dg\_prep\_launcher2\_01.py script in each of the MF-FEB directories to compute FEB. The dg\_prep\_launcher2\_01.py script also built a master shell script to run the other results scripts. The FEB was written to a text file named fe\_binding.

## **Chapter 3 Optimizing Virtual Screening and Mean Field – Free Energy of Binding Calculations**

Before employing MF-FEB calculations for rescoring or AutoDock4 for virtual screening, we needed to optimize the runtime parameters for both supercomputer and grid computing environments. This chapter describes the parameter selection process, and presents the data that lead to their selection. Section 3.1 describes selecting parameters for AutoDock4 for virtual screening and processing virtual screening results. Section 3.2 describes preparing the MF-FEB calculations for rescoring virtual screening results.

### **3.1 VIRTUAL SCREENING**

This section discusses the selection of parameters for AutoDock4 and the processing of docking results for virtual screening on World Community Grid. We used the same parameters for virtual screening projects using TACC high performance computing resources as we did on World Community Grid. Section 3.1.1 describes the optimization of single docking simulations for virtual screening. Section 3.1.2 describes processing virtual screening results to select compounds for rescoring using MF-FEB calculations.

#### **3.1.1 Docking Parameters for Virtual Screening**

In order to use AutoDock4 effectively for virtual screening, we needed to address two questions. The first question was which of all the poses generated by AutoDock4 was closest to the experimental pose, as we were only rescoring one pose from a docking simulation due to limited computational resources? The second question was what were the ideal search parameters for a virtual screening with AutoDock4, balancing runtime with an effective search of the binding energy landscape? The goal of this section is to show how we arrived at the solutions to these questions. Section 3.1.1.1 introduces the trypsin test set used to answer these

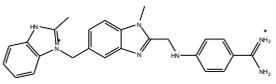
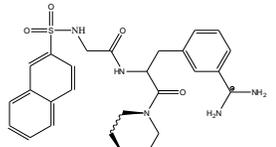
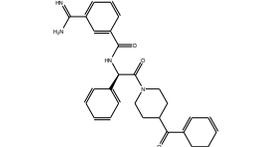
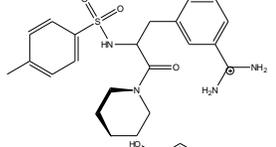
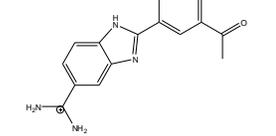
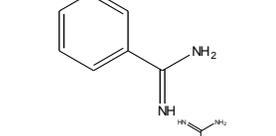
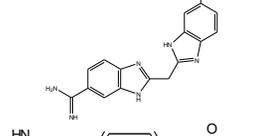
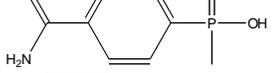
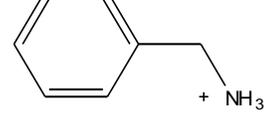
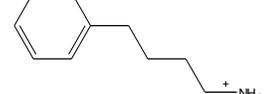
questions. Section 3.1.1.2 discusses result pose selection. Section 3.1.1.3 discusses the selection of search parameters for a virtual screening with AutoDock4.

### ***3.1.1.1 Trypsin Test Set***

As the dengue NS3 protease was our primary target for DDDT, we selected the well-studied trypsin protease to serve as a test system to optimize AutoDock4 for virtual screening. Trypsin is similar to the dengue NS3 protease, as both proteins are serine protease cutting after di-basic residues. However, trypsin's binding site is more contoured and charged than dengue's shallower binding site. The training set, composed of ten trypsin inhibitors, was taken from the literature. Each inhibitor had an experimentally determined pose, co-crystal, and experimentally determined  $K_i$  or FEB. Table 3-1 contains a detailed list of inhibitors, including their structures, sources, and the FEB to trypsin. The selected inhibitors covered a range of FEBs, between  $-10 \text{ kcal mol}^{-1}$  and  $-2 \text{ kcal mol}^{-1}$ , and a range of rotatable bonds, between 1 to 7.

The test set had two limitations. First, 8 of the 10 inhibitors contained benzyl-diamine, which binds tightly to the P1 site in trypsin. The remaining two were benzyl-monoamines, which also bind to the P1 site. Therefore, the compounds in the test set were chemically similar, limiting the chemical space defined by the test set. Fortunately, the "tail" portions of the inhibitors, the part outside of the P1 pocket that interacts with other portions of trypsin, varied, increasing the chemical space defined by the test set. The second limitation of the test set was that the  $K_i$ s were obtained from a variety of different sources; therefore, experimental methods were not consistent. Ideally, the properties of the compounds in a test set would be measured using a consistent method in the same lab, but when the test set was compiled there was no single source that contained the same chemical variety as the test set from a single lab. In most cases, the sources gave only a  $K_i$  for the compounds which we then converted to a FEB at 25°C.

Table 3-1 – The Trypsin Test Set (\*Number of bonds allowed to rotate in AutoDock4 docking simulations)

| PDB  | $\Delta G$<br>kcal/mol | Number of<br>Rotatable<br>bonds<br>(longest<br>segment)* | Ligand   | Ligand Structure  |
|------|------------------------|--|--|---|
| 1G36 | -10.2                  | 6(2)   | 4-[[1-METHYL-5-(2-METHYL-BENZOIMIDAZOL-1-YLMETHYL)-1H-BENZOIMIDAZOL-2-YLMETHYL]-AMINO]]BENZAMIDINE |    |
| 1PPC | -8.4                   | 9(3)   | N $\alpha$ -(2-NAPHTHYL-SULPHONYL-GLYCYL)-DL-P-AMIDINOPHENYLALANYL-PIPERIDINE (NAPAP)              |    |
| 1EB2 | -8.2                   | 7(2)   | 3-[(Z)-AMINO(IMINO)METHYL]-N-[2-(4-BENZOYL-1-PIPERIDINYL)-2-OXO-1-PHENYLETHYL]BENZAMIDE            |    |
| 1PPH | -8.1                   | 7(3)   | P-TOLUENE SULFONATE, M-AMIDINOPHENYLALANYL GROUP, PIPERDINE  |   |
| 1GI5 | -7.6                   | 3(1)   | 2-(2-HYDROXY-5-METHOXY-PHENYL)-1H-BENZOIMIDAZOLE-5-CARBOXAMIDINE                                   |  |
| 1BTY | -7.3                   | 1(1)   | BENZYLDIAMINE  |  |
| 1XUI | -6.4                   | 4(2)   | BIS(5-AMIDINO-2-BENZIMIDAZOLYL)METHANONE   |  |
| 1TX7 | -6.3                   | 2(1)   | (4-CARBAMIMIDOYLPHENYL)-METHYL-PHOSPHINIC ACID   |  |
| 1UTN | -4.7                   | 2(2)   | BENZYLAMINE  |  |
| 1UTP | -2.0                   | 5(4)   | 4-PHENYLBUTYLAMINE   |  |

### ***3.1.1.2 Pose Selection***

As discussed in Section 1.2.3, docking programs can accurately reproduce experimentally determined binding poses. (Kellenberger, Rodrigo et al. 2004; Chen, Lyne et al. 2006; Warren, Andrews et al. 2006; Cross, Thompson et al. 2009) However, AutoDock4 provided anywhere from 10 to 255 poses for the docking simulation of one ligand to one target due to repeated GA runs. AutoDock4 then subjected resulting poses to cluster analysis, as described in Section 2.3.2.3. The cluster analysis provided two criteria for each pose: score and cluster size. The challenge was to select systematically the “best” pose from all the docking results because we would be using only one pose for MF-FEB calculations. The “best” pose was the pose most similar to the experimental, or co-crystal, pose as determined using RMDS. To solve the problem of pose selection, we performed docking simulations on the trypsin test set using different docking parameters and target conformations. We determined that selecting the lowest energy pose of the largest cluster was the most effective method for systematically selecting the best pose from AutoDock4’s results. This section explains how we arrived at that conclusion.

To determine a method for systematically selecting the best pose from the docking results, we ran two sets of docking simulations using the trypsin test set. We employed the same search volume for all docking simulations. The search volume was sufficiently large to include all ligands in the test set and corresponded in size to a virtual screening search volume as described in Section 2.3.1.2. In the first simulations, we self-docked (docked into their own trypsin structure) all ten ligands, and then cross docked the test set into the trypsin structure 1EB2. We ran these docking simulations 200 times (i.e., 200 GA runs) for each ligand with a MNEE of 250,000. In the second simulation set, we self-docked all ten ligands at 200 GA runs, running each docking at a MNEE of 250,000 and 1 million.

Our first task was to determine which method using cluster size and score would select the “best” pose. Because a cluster is composed of multiple poses, we selected the lowest scoring pose, which theoretically corresponds to the lowest energy pose, as the representative pose for a

cluster, and identified the pose by its score and the size of the cluster it represented. Generally, the lowest scoring pose in a cluster approximates the experimental pose best by having the lowest RMSD relative to the experimental pose. We examined four results selection criteria: first, the lowest scoring pose; second, the pose of the lowest scoring cluster; third, the pose of the lowest scoring cluster that also contained at least 5% of the results; and fourth, the largest cluster (the lower scoring of multiple matching cluster sizes). The results of the different docking simulations are shown in Table 3-2.

With all the selection criteria, the average RMSD was  $\sim 1$  Å, slightly better than the general performance of docking programs (Kellenberger, Rodrigo et al. 2004; Chen, Lyne et al. 2006; Warren, Andrews et al. 2006; Cross, Thompson et al. 2009); however, no selection criteria performed significantly better than any other selection criteria. AutoDock4 was able to produce a pose with an average RMSD of 0.5 Å for all ligands; however, no selection criteria selected the “best” pose. In addition, on average, there was no significant difference between self and cross-docked poses, demonstrating the relatively static nature of the trypsin binding site. We self-docked 1EB2 in both simulation sets and showed a difference of 0.1 Å between docking simulation results.

As expected, an increase in ligand flexibility correlated to an increase in RMSD. Neither 1pph nor 1ppc, which are both part of the same synthetic series and the most flexible inhibitors, generated any clusters, suggesting that the MNEE was insufficient to generate convergent results. Encouragingly, the lowest energy poses for 1PPH and 1PPC had RMSDs less than 1.5 Å.

While no single selection criteria performed significantly better than any other, the best average performer was “largest cluster”. Lowest scoring cluster selection criteria shared poses with either the largest cluster or lowest score; therefore, neither provided additional discrimination. Interestingly, the lowest scored pose as defined by the scoring function, which

should be the nearest to the energetic minimum for the binding energy landscape, did not reproduce the best pose, and in the case of 1G36 selected a pose significantly different from the co-crystal pose.

The selection criterion established by these results in Table 3-2 was to take the lowest energy pose of the largest cluster, selecting the lowest scoring cluster if multiple clusters were the same size, or the lowest scoring pose if no clusters were present. We refer to this selection criterion as the “standard pose selection method.”

In the previous docking simulations, neither 1pph nor 1ppc formed any clusters; therefore, we performed a second set of docking simulations increasing the MNEE from 250,000 to 1 million to determine the impact on our result selection criterion. The results of the second set of simulations are presented in Table 3-3. The increase in MNEE generally decreased the number of resulting clusters.<sup>11</sup> The number of clusters correlates ( $r = 0.7$ ) to the number of rotatable bonds in each ligand. Additionally, both 1PPH and 1PPC generated clusters at an MNEE of 1 million. The average of the best overall poses was again 0.5 Å, which is consistent with the previous simulations.

While largely consistent, variation in the 250,000 best pose column compared to Table 3-1 was due to variation in the random number of seeds for each docking run, with differences between the tables being less than 0.1 Å. At a MNEE of 250,000, there was no significant difference between the standard pose selection method and the lowest scoring pose, though the average RMSD of the standard pose selection method was less. At an MNEE of 1 million, there was a significant difference between the two selection criteria ( $p = 0.04$ ). This increase in significance was caused by an increase in RMSD of the lowest scoring poses as the MNEE was increased. An insignificant decrease in RMSD was seen using the standard pose selection

---

<sup>11</sup> In these results, 200 clusters corresponded to no clusters formed, or 200 clusters of one member each.

method. These results supported the effectiveness of the standard pose selection method and suggested that a higher MNEE was required for good docking results.

For AutoDock4, the standard pose selection method was shown to be the best method for systematically selecting poses from the results that best matched experimental poses. Interestingly, the lowest scoring poses did not correspond to the experimental poses. Additionally, with the increase in MNEE, the RMSD of the lowest scoring poses increased. These two observations show that the lowest minimum in the binding energy landscape, as reported by the scoring function, did not correspond to the experimental binding pose. However, in the cases where the RMSD was greater than 2 Å the average difference in score was less than 0.2 kcal mol<sup>-1</sup>. Additionally, with 1GI5 and 1XUI, their lowest scoring poses varied less than 1Å while the average score decreased by greater than 1 kcal mol<sup>-1</sup>. It is possible the AutoDock4 was finding multiple natural binding conformations not seen in the co-crystal structures, however, with the reported error of the scoring function being ~2 kcal mol<sup>-1</sup> (Huey, Morris et al. 2007), it is unlikely that the scoring function is sensitive enough to confidently accept this conclusion. More likely, the RMSD divergence was due to inaccuracies in AutoDock4's scoring function approximation of the binding energy landscape. AutoDock4's scoring function is largely an enthalpic measurement focused on interaction energies. It is possible that the global minimum in the enthalpic energy landscape do not correspond to the global minimum in the binding energy landscape. Fortunately, the overall approximation of the landscape was sufficient that the consensus score was most often a good binding pose.

Table 3-2 – The RMSD (Å) of AutoDock4 generated poses, selected using a variety of criteria, to the experimentally determined poses.

| System<br>pdb ID | Number<br>of<br>Rotatable<br>Bonds in<br>docking | Largest<br>Number<br>of<br>Continues<br>of<br>Bonds | Best RMSD    |                   | Best Score   |                   | Top Scoring<br>Cluster |                   | Top Scoring<br>Cluster With at<br>Least 5% of<br>Compounds |                   | Largest Cluster |                   |
|------------------|--|---|--------------|-------------------|--------------|-------------------|------------------------|-------------------|--|-------------------|-----------------|-------------------|
|                  |  |   | Self<br>Dock | Docked<br>to 1eb2 | Self<br>Dock | Docked<br>to 1eb2 | Self<br>Dock           | Docked<br>to 1eb2 | Self<br>Dock   | Docked<br>to 1eb2 | Self<br>Dock    | Docked<br>to 1eb2 |
| 1BTY             | 1  | 1   | 0.20         | 0.14              | 0.25         | 0.55              | 0.28                   | 0.26              | 0.28   | 0.26              | 0.28            | 0.26              |
| 1TX7             | 2  | 1   | 0.14         | 0.40              | 0.53         | 0.77              | 0.35                   | 0.73              | 0.35   | 0.73              | 0.35            | 0.73              |
| 1UTN             | 2  | 2   | 0.61         | 0.45              | 0.70         | 0.64              | 0.73                   | 0.62              | 0.73   | 0.62              | 0.73            | 0.62              |
| 1GI5             | 3  | 1   | 0.42         | 0.41              | 1.37         | 0.79              | 1.72                   | 0.78              | 0.56   | 1.54              | 0.56            | 0.54              |
| 1XUI             | 4  | 2   | 0.42         | 0.41              | 1.37         | 0.79              | 1.72                   | 0.78              | 0.56   | 1.54              | 0.56            | 0.54              |
| 1UTP             | 5  | 4   | 1.03         | 0.95              | 1.67         | 2.83              | 1.64                   | 3.11              | 1.27   | 1.80              | 1.27            | 1.80              |
| 1G36             | 6  | 2   | 0.41         | 0.50              | 6.63         | 1.00              | 6.56                   | 0.94              | 0.76   | 0.89              | 0.76            | 1.96              |
| 1EB2             | 7  | 2   | 0.50         | 0.40              | 0.82         | 0.82              | 0.75                   | 0.78              | 0.75   | 0.78              | 0.75            | 1.30              |
| 1PPH             | 7  | 3   | 0.84         | 0.86              | 1.43         | 0.86              | NA                     | NA                | NA   | NA                | NA              | NA                |
| 1PPC             | 9  | 3   | 0.71         | 0.85              | 0.71         | 1.37              | NA                     | NA                | NA   | NA                | NA              | NA                |
| Average          |  |   | 0.53         | 0.54              | 1.55         | 1.04              | 1.72                   | 1.00              | 0.66   | 1.02              | 0.66            | 0.97              |

Table 3-3 – Changes in RMSD between docking generated and experimentally determined ligand pose at different maximum number energy evaluations parameter settings.

| System<br>pdb ID | Number<br>of<br>Rotatable<br>Bonds in<br>docking | Largest<br>Number<br>of<br>Continues<br>Bonds | Number of<br>Clusters |     | Lowest Score<br>Pose<br>RMSD (Å) |      | Largest<br>Cluster Pose<br>RMSD (Å) |      | Best Pose<br>RMSD (Å) |      |
|------------------|--|---|-----------------------|-----|----------------------------------|------|-------------------------------------|------|-----------------------|------|
|                  |  |   | 250k                  | 1m  | 250k                             | 1m   | 250k                                | 1m   | 250k                  | 1m   |
|                  |  |   | 1BTY                  | 1   | 1                                | 1    | 1                                   | 0.25 | 0.28                  | 0.25 |
| 1TX7             | 2  | 1   | 2                     | 2   | 0.53                             | 0.59 | 0.53                                | 0.59 | 0.14                  | 0.18 |
| 1UTN             | 2  | 2   | 2                     | 2   | 0.70                             | 0.78 | 0.70                                | 0.78 | 0.61                  | 0.69 |
| 1GI5             | 3  | 1   | 92                    | 33  | 0.93                             | 1.05 | 0.52                                | 0.45 | 0.34                  | 0.40 |
| 1XUI             | 4  | 2   | 118                   | 60  | 1.37                             | 1.36 | 0.64                                | 0.87 | 0.42                  | 0.47 |
| 1UTP             | 5  | 4   | 120                   | 43  | 1.67                             | 3.29 | 1.77                                | 1.57 | 1.01                  | 1.23 |
| 1G36             | 6  | 2   | 58                    | 17  | 6.63                             | 6.65 | 0.85                                | 0.91 | 0.41                  | 0.59 |
| 1EB2             | 7  | 2   | 73                    | 10  | 0.82                             | 0.92 | 1.31                                | 0.92 | 0.50                  | 0.37 |
| 1PPH             | 7  | 3   | 200                   | 158 | 1.43                             | 5.23 | 1.43                                | 0.58 | 0.84                  | 0.35 |
| 1PPC             | 9  | 3   | 200                   | 185 | 0.71                             | 3.94 | 0.71                                | 0.39 | 0.71                  | 0.39 |
| Average          |  |   |                       |     | 1.50                             | 2.41 | 0.87                                | 0.73 | 0.52                  | 0.49 |

### ***3.1.1.3 AutoDock4 Search Parameters for Virtual Screening***

Ideally, virtual screening parameters would exhaustively explore the binding energy landscape for every possible ligand-target combination. However, the parameters to perform an exhaustive search in AutoDock4 demand significant amounts of time making virtual screening onerous, and negating the advantage of a stochastic search method. Therefore, we needed to determine a set of search parameters that effectively explore the binding energy landscape in a tractable amount of time. Because we intended to use these virtual screening parameters on World Community Grid, we targeted a runtime of ~2 hours for each docking simulation. We performed two tests to determine the optimal search parameters of AutoDock4. The key parameters that determine the runtime for a docking simulation were the number of docking simulations pre-ligand, or genetic algorithm runs (GA runs), and the max number energy evaluations (MNEE). We first screened multiple combinations for GA runs and MNEEs selecting 100 runs and 1million MNEE as the ideal parameters, known herein as production parameters. Secondly, the poses generated using production parameters were compared with pose generated against a “gold” standard parameter set to determine the quality of pose generation. We show that there was no significant difference in pose quality between the two parameter sets. This section presents the methods, data, and a discussion of these two tests.

In order to find the optimal search parameters, we self-docked the members trypsin test set using a variety of MNEE from 100,000 to 2 million in five half log steps, and GA runs from 10 to 200 in five ~50 run increments. To avoid running all 25 combinations, we modified only one parameter at a time for a total of nine runs. As an MNEE of 1 million and 100 GA runs was the midpoint of the two ranges, they were selected as the constant values while we varied the other parameter (i.e., MNEE runs were varied, but GA runs were kept at 100). The results of the screen are presented in Table 3-4. The average runtime of all compounds for both parameters scaled linearly, so it was reasonable to assume the overall surface generated by varying both parameters simultaneously would be a plain. Conveniently, the center of the parameter screen,

MNEE of 1 million and 100 GA runs, also corresponded to our timing goal of ~2 hours at about  $1.5 \pm 0.5$  hours. Therefore, we selected 100 GA runs and a MNEE of 1 million as our production standard. Consistent with previous work, the RMSD of docking poses did not significantly vary with the changes in parameters. However, the trends indicate that at least 50 GA runs will produce a better pose results than 10, as the difference in the averages is greater than the 0.10 Å difference caused by the random number error. Additionally, the RMSD tends to increase with the increase in MNEE as discussed in Section 3.1.1.2.

As an additional evaluation, we compared the pose and timing of self-docking the trypsin test set using the production standards to those of a “gold” standard, performing each docking simulation 3 times with different random number seeds. The gold standard was a MNEE of 10 million and 200 GA runs, or the max values of the ranges screened above. The gold standard is short of AutoDock4’s maxim whose GA runs are limited to 255 and MNEE to at least 4.2 billion.<sup>12</sup> However, with predicted runtime reaching 40 hours, the gold standard represented the maxim reasonably testable runtime. The times and pose RMSD for each ligand in the trypsin test set, using both the gold and production parameter sets, is shown in Table 3-5.

There was a significant difference in average run time between the gold standard, at ~30 hours, and the production standard, at ~2 hours. The gold standard timing was consistent with the linear increase in time seen in the previous test, and strengthens the plainer assumption in varying both GA runs and MNEE. As expected, the runtime for each ligand generally increased with the number of rotatable bonds. Most importantly, there was no significant difference between the average RMSD for the selected poses between the production and gold parameters, and therefore no apparent advantage was gained from the more time consuming virtual screening parameters. This supports the selection of the production parameters for virtual screening experiments.

---

<sup>12</sup> Assuming the variable was assigned as a long integer.

The trypsin test set has ligands with varying flexibility that provide a range of docking simulations of different complexity requiring a varied amount of runtime. Additionally, the compounds in the trypsin test set are drug-like. Therefore, we could reasonably apply the production parameters to other compound databases, and expect the similar runtimes and pose quality.

Table 3-4 – Average RMSD of AutoDock4 generated pose vs. experimental determined poses and average runtime for the trypsin test set using different docking parameters.

| Parameter                    | Avg. RMSD to Crystal<br>(Å) | Avg. Timing<br>(Hours) |
|------------------------------|-----------------------------|------------------------|
| Number of Runs               |                             |                        |
| 200                          | 1.51 ± 0.87                 | 2.91 ± 1.11            |
| 150                          | 1.45 ± 0.71                 | 2.17 ± 0.82            |
| 100                          | 1.45 ± 0.69                 | 1.45 ± 0.55            |
| 50                           | 1.46 ± 0.64                 | 0.73 ± 0.28            |
| 10                           | 1.79 ± 0.65                 | 0.15 ± 0.06            |
| Number of Energy Evaluations |                             |                        |
| 10,000,000                   | 1.78 ± 1.18                 | 14.56 ± 5.54           |
| 5,000,000                    | 1.72 ± 1.17                 | 7.27 ± 2.77            |
| 1,000,000                    | 1.45 ± 0.69                 | 1.45 ± 0.55            |
| 500,000                      | 1.08 ± 0.32                 | 0.73 ± 0.27            |
| 100,000                      | 1.39 ± 0.50                 | 0.15 ± 0.06            |

Table 3-5 – Comparison of “Gold” and “Production” docking parameters average RMSD of AutoDock4 generated pose vs. experimental determined poses and average runtime for each compound in the trypsin test set. n=3

| Ligand (pdb ID) | Number of Rotatable Bonds in docking | Largest Number of Continuous Bonds | RMSD to Crystal (Å) |             | Timing (hours) |             |
|-----------------|--------------------------------------|------------------------------------|---------------------|-------------|----------------|-------------|
|                 |                                      |                                    | Gold                | Production  | Gold           | Production  |
| 1BTY            | 1                                    | 1                                  | 0.56 ± 0.00         | 0.56 ± 0.00 | 9.14 ± 0.36    | 0.45 ± 0.01 |
| 1TX7            | 2                                    | 2                                  | 1.22 ± 0.30         | 0.91 ± 0.32 | 9.27 ± 0.32    | 0.45 ± 0.01 |
| 1UTN            | 2                                    | 1                                  | 0.79 ± 0.00         | 0.80 ± 0.01 | 12.18 ± 0.09   | 0.64 ± 0.00 |
| 1GI5            | 3                                    | 1                                  | 1.75 ± 0.03         | 1.68 ± 0.10 | 19.06 ± 0.00   | 1.15 ± 0.02 |
| 1XUI            | 4                                    | 2                                  | 1.33 ± 0.02         | 1.32 ± 0.04 | 36.61 ± 0.33   | 1.81 ± 0.02 |
| 1UTP            | 5                                    | 4                                  | 3.31 ± 0.17         | 3.27 ± 0.10 | 12.93 ± 0.16   | 0.64 ± 0.01 |
| 1G36            | 6                                    | 2                                  | 6.65 ± 0.01         | 3.66 ± 2.49 | 42.79 ± 0.54   | 2.14 ± 0.01 |
| 1EB2            | 7                                    | 2                                  | 0.86 ± 0.01         | 0.88 ± 0.00 | 47.55 ± 0.44   | 2.36 ± 0.04 |
| 1PPH            | 7                                    | 3                                  | 0.62 ± 0.07         | 0.45 ± 0.07 | 40.20 ± 0.58   | 1.99 ± 0.03 |
| 1PPC            | 9                                    | 3                                  | 0.47 ± 0.15         | 0.99 ± 0.88 | 57.14 ± 0.34   | 2.85 ± 0.01 |

### **3.1.2 Filtering Virtual Screening Results for MF-FEB Rescoring on World Community Grid**

Due to the time requirement of the free energy of binding calculations, rescoring all 2.3 million compounds used for the AutoDock4 virtual screening was impractical even with the computational resources of World Community Grid.<sup>13</sup> However, rescoring a few thousand compounds, which represents a tenth of one percent of the whole database, was a practical option. We therefore needed to establish a filtering method to select which compounds from AutoDock4's virtual screening to rescore using the MF-FEB calculations. As discussed in Section 1.3.2, virtual screening can enrich a compound database for binders. Logically, we could have selected the lowest, "best" scoring compounds, however, AutoDock4's results, processed using the standard pose selection method, have both a score and a clusters size. Additionally, as discussed in Section 3.1.1.2, cluster size was more important to pose selection than score. Therefore, we asked if there was a combination of pose, cluster size, or both that would increase the probability of selecting true positives from the AutoDock4 virtual screening for rescoring using MF-FEB calculation.

Using five targets from the DUD dataset we determined that best approach for compound selection was to take the first 2,000 lowest scoring compounds and the 2,000 largest cluster size compounds. This section describes how we reached this conclusion. Section 3.1.2.1 describes the test sets and methods used to select a filtering method. Section 3.1.2.2 discusses the results of five virtual screening experiments and results using the filtering method.

#### ***3.1.2.1 Methods and Test Set***

To determine the optimal filtering method for selecting which compounds to rescore, we selected five test sets from the DUD database. (Huang, Shoichet et al. 2006) We selected: (1) estrogen receptor agonists (ERAG); (2) HIV-1 protease (HIVP); (3) HIV-1 reverse transcriptase

---

<sup>13</sup> See Chapter 5 and Chapter 6 for a discussion on our World Community Grid projects and grid implementation of AutoDock4.

(HIVR); (4) neuraminidase (NADU); and (5) trypsin (TRYP). Table 3-6 contains the number of binders and decoys for each system, and the target pdb ID for each test set. We selected targets that represented a variety of binding site architectures. ERAG was selected because the binding site is not solvent exposed and the intermolecular interactions between the ligands protein are largely VdW interactions with a few key H-bonds. We chose the agonist system over the antagonists due to its better performance and enrichment. (Huang, Shoichet et al. 2006; Cross, Thompson et al. 2009) We selected HIVR because like ERAG its binding site was solvent excluded, but intermolecular interactions of binding were more dependent on electrostatic interactions and less on VdW interactions than ERAG. HIVP was selected because the binding site is slightly solvent exposed though still largely defined by the protein. TRYP and NADU were selected because of their use in DDDT and IADS, respectively. Additionally, these systems' binding sites were solvent exposed, and their bindings were defined largely by electrostatic interactions. Therefore, each system would increasingly deviate from the L99A T4 lysozyme system we used to originally test the MF-FEB calculations (See Chapter 4).

Table 3-6 – Selected DUD targets.

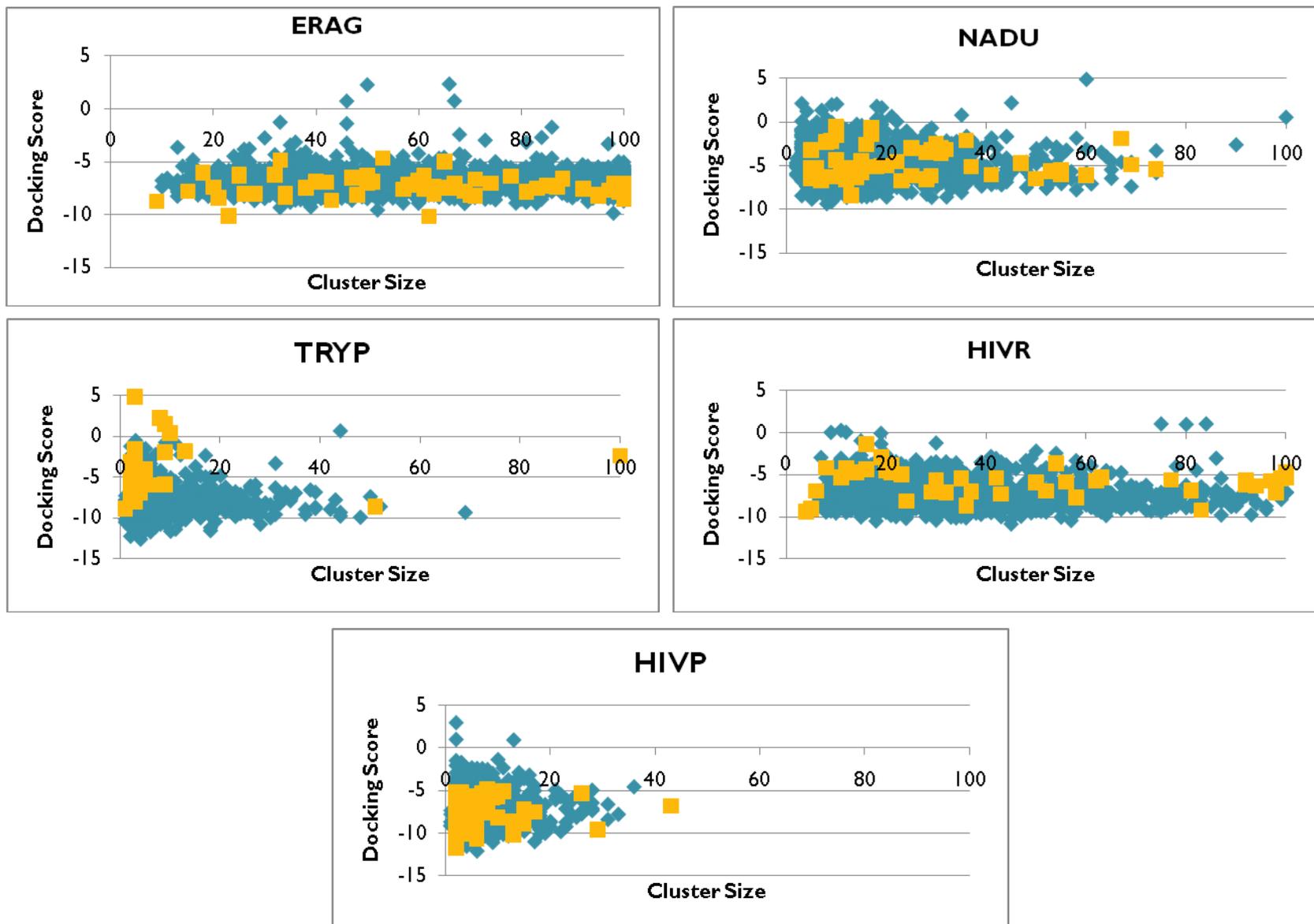
| DUD System                   | Number of Binders | Number of Decoys | Target Crystal Structure |
|------------------------------|-------------------|------------------|--------------------------|
| Estrogen Receptor (Agonists) | 67                | 2,361            | 1L21                     |
| HIV-1 Reverse Transcriptase  | 40                | 1,439            | 1RT1                     |
| HIV-1 Protease               | 53                | 1,888            | 1HPX                     |
| Neuraminidase                | 49                | 1,745            | 1A4G                     |
| Trypsin                      | 43                | 1,545            | 1BJU                     |

We performed compound database preparation, target preparation, virtual screening, and results collection as described in Section 2.3.2. The compound database for each virtual screening included the co-crystal ligand preparation as described in Section 2.2.1.1.1 and 2.2.1.2.

### ***3.1.2.2 Results and Discussion***

To identify a result filter, we plotted the docking results from each virtual screening experiment on a scatter plot with the docking score on the y-axis and the cluster size on the x-axis, with binders and decoys differently colored to visualize their distribution. Figure 3-1 contains the scatter plots for all five DUD test sets. Table 3-7 contains the averages and the 95% confidence intervals for the binders and decoys for all five test sets. Additionally, Table 3-7 contains the p-score from a Student t-test between the binder and decoy populations. In general, as can be seen when comparing the scatter plot in Figure 3-1 and the summary statistics in Table 3-7, the distribution of binders relative to decoys was different from test set to test set. In the ERAG test set, we saw significant discrimination using both score and cluster size, with binders scoring lower and having larger cluster sizes. In the HIVR test set, there was a significant discrimination in both score and cluster size. However, unlike ERAG, the binders scored higher than the decoys, making the docking score an inverse indicator of binding. In the HIVP test set, neither score nor cluster size significantly discriminated between the binders and the decoys. For the NADU test set, only cluster size significantly discriminated between binders and decoys. Finally, with the TRYP test set, only cluster size significantly discriminated between binders and decoys, but inversely.

Figure 3-1 – Scatter plots of docking results by docking score and cluster size for five DUD test sets. Yellow squares are binders and blue diamonds are decoys.



The enrichment rates' dependence on system was consistent with previous virtual screening performance studies. (Kellenberger, Rodrigo et al. 2004; Kontoyianni, McClellan et al. 2004; Warren, Andrews et al. 2006; Cross, Thompson et al. 2009) Overall, AutoDock4 underperformed for these five DUD test sets when compared to other docking programs. (Cross, Thompson et al. 2009) However, the question of filtering criteria remained. In reviewing the discrimination data, neither docking score nor cluster size consistently discriminated the binders from the decoys. Additionally, in cases where a result metric population was statistically different, populations still overlapped producing the false positives results. However, we expected the large number of false positives, hence the MF-FEB rescoring.

By examining the scatter plots, we observed that as one selects compounds by rank order, by either cluster size or score, binders would be included in the first few percent of a least one of the two results metrics. For example, while rank ordering the TRYP test set by score, taking the first few percent of the lowest scores would net no binders, but the cluster metric would. In the case of ERAG, the first few percent of each metric would select a binder. Of the two metrics, cluster size was the most likely to find binders, which was consistent with the standard pose selection method. Additionally, the selection by one metric at a time was more likely to select binders than the lowest scoring and largest cluster results.

We decided to filter the results by taking the first 2000 compounds ranked by cluster size and by score; neither list containing the same compounds. The advantage of this approach was that we were more likely to get binders in the compounds that we could rescore. However, in some cases the number of binders might be relatively small as compared to the number of compounds rescored, meaning that MF-FEB rescoring would need to be much more effective at discriminating binders from non-binders than AutoDock4's scoring function, to find the binders. Additionally, we observed that in virtual screening some of the lowest and highest scoring compounds were often erroneous predictions due to errors in the ligand structures caused by

automatic database generation. Therefore, a number of the compounds selected for rescoring based on docking score would likely be “bad” and cause errors in the MF-FEB calculations.

Table 3-7 – Population statistics for score and clusters size for each DUD virtual screening.

| DUD System                   | Result Metric                   | Binders      | Decoys       | T-Test (p-score) |
|------------------------------|---------------------------------|--------------|--------------|------------------|
| Estrogen Receptor (Agonists) | Score (kcal mol <sup>-1</sup> ) | -7.42 ± 0.23 | -6.74 ± 0.03 | 0.0000003        |
|                              | Cluster size                    | 65 ± 7       | 57 ± 1       | 0.03             |
| HIV-1 Reverse Transcriptase  | Score (kcal mol <sup>-1</sup> ) | -5.85 ± 0.49 | -7.32 ± 0.09 | 0.0000006        |
|                              | Cluster size                    | 47 ± 10      | 33.78 ± 1    | 0.01             |
| HIV-1 Protease               | Score (kcal mol <sup>-1</sup> ) | -7.96 ± 0.40 | -7.57 ± 0.07 | 0.06             |
|                              | Cluster size                    | 6 ± 2        | 5.2 ± 0.2    | 0.1              |
| Neuraminidase                | Score (kcal mol <sup>-1</sup> ) | -4.54 ± 0.51 | -4.59 ± 0.08 | 0.8              |
|                              | Cluster size                    | 26 ± 5       | 14.4 ± 0.5   | 0.00004          |
| Trypsin                      | Score (kcal mol <sup>-1</sup> ) | -4.91 ± 0.80 | -7.75 ± 0.08 | 0.0000001        |
|                              | Cluster size                    | 7 ± 4        | 6.5 ± 03     | 0.9              |

## **3.2 MF-FEB CALCULATIONS**

To utilize the MF-FEB calculations, we optimized them for run time length and FEB prediction accuracy. Section 3.2.1 describes the optimization of the system size and microstate sampling. Section 3.2.2 describes the optimization of the GCMC simulations in the site set-up stage of the MF-FEB calculations.

### **3.2.1 Local Run MF-FEB Optimization**

Before utilizing MF-FEB for rescoring, the calculation parameters needed to be optimized to provide accurate FEB prediction; however, with a total runtime of ~550 hours for a single MF-FEB calculation, only a limited number of parameters could be effectively optimized. Conveniently, Deng et al. had already optimized many of the MF-FEB calculation parameters. (Deng and Roux 2006; Wang, Deng et al. 2006) Therefore, we asked three questions the previous works did not address. First, what was the optimal size for the solvent drop? Second, were the perturbation MD run lengths sufficient for convergence? Third, what was the variance in the energy contribution of each perturbation stage between perturbation runs started with different seed values for the random number generator? This section explores the answer to these three questions. Section 3.2.1.1 describes the methods used to answer the questions. Section 3.2.1.2 discusses optimal system size for MF-FEB calculations. Section 3.2.1.3 discusses the question of convergence. Section 3.2.1.4 discusses the variation in energy due to different random number seeds.

#### ***3.2.1.1 Methods***

The binding of hydrophobic small molecules to the engineered binding pocket of mutant L99A T4 lysozyme provides an ideal test system, as the experimental FEB were determined by one lab, and the developers of the MF-FEB method also utilized the same test system. (Morton, Baase et al. 1995; Morton and Matthews 1995; Deng and Roux 2006) To answer our three questions, we calculated the free energy of binding benzene to the L99A T4 lysozyme mutant. The experimentally determined FEB of benzene to the engineered hydrophobic pocket of the

L99A T4 lysozyme mutant was  $-5.2 \pm 0.2 \text{ kcal mol}^{-1}$ . (Morton, Baase et al. 1995) We obtained the benzene-lysozyme co-crystal coordinates from Dr. Y. Deng (University of Chicago). We performed the set-up stage for both the site and bulk systems for each ligand according to the methods described in Section 2.4.2, using the first generation scripts provided by Drs. Roux and Deng. We ran the perturbation calculation four times using the same set-up stage results but varying the random number seed, and reported FEB as the average of the four runs with a 95% confidence interval.

### 3.2.1.2 System Size

Originally, Deng et al. selected a solvent drop size of  $15 \text{ \AA}$ , using a non-bonding cutoff of  $12 \text{ \AA}$ , and calculated the FEB as  $-5.96 \pm 0.19$  ( $n=3$ ), which was  $\sim 1 \text{ kcal mol}^{-1}$  too low. (Deng and Roux 2006) In order to determine the effect of drop size on the FEB, we varied the drop size to 12, 15 and  $18 \text{ \AA}$ , maintaining the  $12 \text{ \AA}$  non-bonding cutoff distance. Table 3-8 contains the results of those calculations.

Table 3-8 – Comparison of free energy of binding calculations using MF-FEB for benzene binding to L99A T4 lysozyme with different solvent drop radii.  $n=4$

| Solvent Drop Radius ( $\text{\AA}$ ) | $\Delta\Delta G_{\text{binding}}$ ( $\text{kcal mol}^{-1}$ ) |
|--------------------------------------|--|
| 12                                   | $-3.58 \pm 0.50$   |
| 15                                   | $-7.33 \pm 1.12$   |
| 18                                   | $-5.01 \pm 0.36$   |

When comparing the FEB predictions of the original work at  $15 \text{ \AA}$ , our FEB calculations were  $\sim 1 \text{ kcal mol}^{-1}$  lower, though not significantly different. This difference may have been due to the differences in the MF-FEB methods, as the previous work did bulk calculations using PBC and not SSBP. The PBC could be the cause of the difference, as the difference between our free energy of solvation and the original was  $\sim 1 \text{ kcal mol}^{-1}$ , which is greater than the originally reported  $\sim 0.1 \text{ kcal mol}^{-1}$ . (Deng and Roux 2006) However, there was also variation in the free

energy from the site calculations of about  $\sim 2$  kcal mol<sup>-1</sup>. The most likely reason for the variation was that we used the GAFF force field for the ligand, whereas in the original study the ligand's force field was cobbled together from parts of the CHARMM force field. (Deng and Roux 2006) Interestingly, when we extended the non-bonding cutoff distance to match the sphere radius of 15 Å, the calculated FEB was  $-4.53 \pm 0.96$  kcal mol<sup>-1</sup>, a better FEB prediction than the original work. When comparing the other sphere sizes to the experimental FEB, the 12 Å underestimates the FEB by  $\sim 1.5$  kcal mol<sup>-1</sup>. However, at an 18 Å sphere radius the prediction for FEB matched the experimental FEB. These results hint at the importance of longer-range interactions in determining the FEB. These interactions were more than 12 Å from the center of the ligand as they were either modeled explicitly or were accurately portrayed in the GSBP at the 15 or 18 Å sphere sizes.

Based on these results, we selected a system size of 15 Å with the non-bond cutoff distance matching to the sphere radius. Unfortunately, later in the work errors arose in the perturbation and GCMC calculations from amino acids moving into the GSBP, which crashed the simulation. We therefore extended the system size to 18 Å, in addition to modifying the GSBP exclusion definitions.

### ***3.2.1.3 Perturbation Sampling Time***

Because FEP methods calculate the FEB based on the sampling of microstates from the MD simulations, we needed to determine if the perturbation steps of the MF-FEB calculations were sufficiently sampling the microstates. We tested sampling efficiency by doubling the number of MD steps in the perturbation MD simulations, which slightly more than doubled the number of sampled states, as we did not extend the starting equilibration phase of the MD simulations. Additionally, we performed the simulation using two different sphere sizes, setting the non-bonding cutoff distance equal to the sphere size. Table 3-9 contains the results of the different runs.

Table 3-9 – Comparison of free energy of binding calculations using MF-FEB for benzene binding to L99A T4 lysozyme with two different solvent drop sizes and two different perturbation MD run times. n=4

| Solvent Drop Radius (Å) | Perturbation MD length (ps) | $\Delta\Delta G_{\text{binding}}$ (kcal mol <sup>-1</sup> ) |
|-------------------------|-----------------------------|---|
| 12                      | 80                          | -3.78 ± 0.50  |
| 12                      | 160                         | -3.79 ± 0.58  |
| 15                      | 80                          | -4.53 ± 0.96  |
| 15                      | 160                         | -4.81 ± 1.03  |

There was no significant difference in FEB due to increased sampling. From the results, we concluded that the FEB had converged within the limitations of the MD simulations to explore conformational space both energetically and within a picoseconds time scale. Although we applied this conclusion of convergence to all MD simulations, we did so with the understanding that as the complexity of the ligand increased, the conclusion's validity would decrease. Therefore, convergence was not guaranteed for all MF-FEB calculations performed on World Community Grid.

#### 3.2.1.4 Variation in Perturbation Stage Free Energies

Throughout the optimization tests above, we observed that the confidence interval for the FEB ranged from 0.36 to 1.03 kcal mol<sup>-1</sup>. These variations were large than originally reported by Deng et al. for the same system. (Deng and Roux 2006) Therefore, we asked the question of where the variation was coming from. Table 3-10 lists the free energy contributions and totals for each system and stage of the MF-FEB calculation for a 15 Å sphere with 15 Å non-bonding cutoff and an 80 ps sampling.

As seen in Table 3-10, most of the variance in FEB came from one simulation. Removing that one simulation would make the FEB -5.09 ± 0.15. For most of the MF-FEB runs in this section, one outlier caused the larger variation. The exception to the trend was the 15 Å sphere, with a 12 Å non-bonding cutoff distance that had two populations, one of -6.16 and -

6.25, which was close to the previously published values, and a second of -8.80 and -8.00. In all cases, the only difference between runs was the random number seed.

As seen in Section 3.2.1.4, the MD simulations were generally convergent and in the individual runs, which were composed the average FEB, the divergent FEB values remained divergent even with increased sampling. The original work varied by only the initial velocities of the MD simulations (Deng and Roux 2006), while in our work the random number seeds were varied, changing not only the MD simulations, but the pre-perturbation minimization, adding an additional degree of divergence between systems. It would be tempting to conclude that the minimization before the MD simulations moved the ligand into a different energy-well that MD simulations could not leave causing the divergent FEB calculations, because when comparing the free energy contributions from the different perturbation steps the greatest variance was in the repulsion and positional constraints energetic contributions. Since the positional constraints were the same for each simulation, the different energy contributions may suggest different minimized poses. Additionally, different repulsion contributions would suggest slightly different steric interactions. Moreover, the difference in energetic contributions of the divergent FEB calculations between repulsion and positional constraints was 6 kcal mol<sup>-1</sup> while the other calculations only differed by 2 kcal mol<sup>-1</sup>. However, to examine the energy-well conclusion fully, we would need to examine the MD trajectory files, which unfortunately were not preserved due to their size.

The general conclusion from the variant results was that one MF-FEB simulation was insufficient to predict confidently the FEB. Therefore, we elected to use multiple MF-FEB runs and determine the FEB as the average of multiple calculations when possible.

Table 3-10 – Comparison of four MF-FEB calculation on benzene binding to L99A T4 lysozyme.

| Simulation System      | Stage                    | Free Energy (kcal mol <sup>-1</sup> ) |        |        |        | Average       | Relative Confidence |
|------------------------|--------------------------|---------------------------------------|--------|--------|--------|---------------|---------------------|
|                        |                          | Run 1                                 | Run 2  | Run 3  | Run 4  |               |                     |
| Site                   | VdW Repulsion            | 6.73                                  | 8.73   | 15.56  | 8.08   | 9.78 ± 3.35   | 34.3%               |
|                        | VdW Dispersion           | -18.97                                | -18.75 | -17.48 | -18.81 | -18.50 ± 0.58 | 3.2%                |
|                        | Electrostatic            | -0.99                                 | -0.81  | -0.76  | -0.92  | -0.87 ± 0.09  | 10.2%               |
|                        | Positional Constraints   | -3.87                                 | -6.28  | -14.51 | -3.18  | -6.96 ± 4.42  | 63.5%               |
|                        | Conformation Constraints | -0.01                                 | -0.01  | -0.01  | -0.01  | -0.01 ± 0.00  | 0.0%                |
|                        | Total Free Energy        | -6.03                                 | -5.61  | -5.92  | -3.64  | -5.30 ± 0.95  | 17.9%               |
| Bulk                   | VdW Repulsion            | 14.30                                 | 14.79  | 14.54  | 14.50  | 14.53 ± 0.17  | 1.2%                |
|                        | VdW Dispersion           | -12.53                                | -12.68 | -12.42 | -12.51 | -12.54 ± 0.09 | 0.7%                |
|                        | Electrostatic            | -2.82                                 | -2.71  | -2.76  | -2.79  | -2.77 ± 0.04  | 1.4%                |
|                        | Total Free Energy        | -1.05                                 | -0.60  | -0.64  | -0.79  | -0.77 ± 0.17  | 22.5%               |
| Free Energy of Binding |                          | -4.98                                 | -5.01  | -5.28  | -2.85  | -4.53 ± 0.96  | 21.1%               |

### **3.2.2 Optimization of GCMC**

The objective of the GCMC stage in the set-up process for MF-FEB calculations is to equilibrate the number of waters in the binding site. Unlike the SSBP, which keeps the solvent drop under constant pressure even if the density of the water molecules changed due to different interactions between the ligand and the bulk solvent, the GSBP was static, and therefore solvent waters may need to be added or removed depending on the ligand in the site. During the optimization and grid implementation, two questions arose that required investigation. First, we observed that the site solvent sphere was losing an abnormal number of waters and therefore we needed to know why this was occurring. Secondly, the runtime for the GCMC simulations was ~50 hours, and as the ideal work unit for World Community Grid was ~10 hours, we wanted to know if all 50 cycles were required.

This section is divided into three parts. Section 3.2.2.1 describes the general methods used in this section. Section 3.2.2.2 answers the first question on water stabilization. Section 3.2.2.3 answers the second question on the required number of cycles.

#### ***3.2.2.1 Methods***

To explore the two questions asked in this section, we used the L99A T4 lysozyme system (See Sections 3.2.1.1 and 4.1) and performed the MF-FEB calculations as described in Section 2.4.

#### ***3.2.2.2 GCMC $\mu$ Scan***

During the site set-up stages for MF-FEB calculations, we equilibrated the number of waters solvating the ligand and the system using a GCMC simulation. Before the GCMC, the waters within the GSBP were either crystallographic waters or waters added from the pre-equilibrated water box as described in Section 2.4.2.1.2. Therefore, at the start of the GCMC simulations the volume occupied by solvent within the GSBP was mostly full. However, we observed that over the course of the GCMC simulations that the total number of waters decreased, leaving a vacuum within the GSBP that GCMC simulation should have filled with

water molecules. In MC simulations, the acceptance or rejection of a move, or in the case of adding molecules in grand canonical simulations, is determined by the chemical potential,  $\mu$ . We concluded that most likely the default  $\mu$  value was incorrect due to the density of water in the binding site, so we tested the effect of different  $\mu$  values on the change in number of waters during GCMC equilibration.

We performed GCMC simulations according to the methods described in Section 2.4.2.1.2, including the MD equilibration, on the benzene-L99A T4 lysozyme system (See Section 3.2.1.1) using four  $\mu$  values of -6.25 (the original value), -4.25, -2.25, and 0 kcal mol<sup>-1</sup>. We ran the GCMC for 25 cycles to ensure that the systems reached equilibrium, and collected the number of waters 10 times per cycle. Figure 3-2 is a plot of the number of waters versus the GCMC cycle. At a  $\mu$  of -6.25 kcal mol<sup>-1</sup>, the number of waters in the GSBP shell decreased until it reached equilibrium at ~120 waters corresponding to the thin layer of water covering the protein and large voids in the solvent volume. At  $\mu$ 's of -4.25, -2.25, and 0 kcal mol<sup>-1</sup> the system quickly reached equilibrium after 10 to 30 cycles, adding 40-50 waters to the system. From the results, it was clear that a  $\mu$  of -6.25 kcal mol<sup>-1</sup> was causing the water to be lost. However, we still needed to know which of the  $\mu$  would correctly equilibrate the waters.

We expected that the GCMC would add or remove only a small number of waters from the site system, as the system began the simulation mostly full of waters added from the equilibrated water box. Therefore, we needed to select a  $\mu$  value that would generate the expected behavior. If we selected a  $\mu$  value that was too low, we would see the number of waters in the system decrease, as was true with a  $\mu$  of -6.25 kcal mol<sup>-1</sup>. If we selected a  $\mu$  that was too high, CHRAMM would add too many waters to the system, increasing the density as seen in the -2.25 and 0 kcal mol<sup>-1</sup> runs. It seemed that a  $\mu$  of -4.25 kcal mol<sup>-1</sup> was “just right,” with the waters equilibrating and freely moving in and out of the sphere over each GCMC/MD cycle, hence the saw-tooth pattern in the graph every 10 check points corresponding to the point of the MD simulations. According to the literature, the appropriate  $\mu$  values for adding water ranges

from -6.25 to -4.25 kcal mol<sup>-1</sup>. (Deng and Roux 2008) We therefore selected a  $\mu$  of -4.25 kcal mol<sup>-1</sup> for our GCMC simulations.

### 3.2.2.3 GCMC Cycles

When implementing MF-FEB in a grid environment,<sup>14</sup> we found that 50 cycles of GCMC were not practical. The GCMC simulations were computationally demanding, thus limiting the number of member devices on the grid able to perform the GCMC simulations. Additionally, the runtime for the GCMC calculations was ~50 hours, while the ideal job size for World Community Grid is 10 hours. The long runtimes and demanding system requirements for the GCMC simulations meant either occupying the highest performance member devices for 50 hours, or dividing the 50 cycles into 10 cycle jobs that would need to be run serially, greatly increasing the total runtime for one MF-FEB calculation on World Community Grid. However, we observed when optimized for a  $\mu$  value of -4.25 kcal mol<sup>-1</sup> the system quickly equilibrated. Therefore, we asked the question of how many cycles the GCMC simulations needed to equilibrate the waters.

To answer the question, we examined the number of waters as a function of cycle numbers for two ligands, a binder and non-binder, for the L99A T4 lysozyme test set described in Section 4.3.1. Figure 3-3 is a plot of the number of waters at the end of each GCMC cycle. For both the binder and the non-binder, the plots show that after the first GCMC cycle the number for waters was within the total range of waters for all 50 cycles, indicating that the number of waters was very close to equilibrium or at equilibrium after only two or three cycles. Additionally, the number of cycles to equilibrium was independent of the experimentally determined binding of the ligand in these two cases. We decided to perform 10 GCMC cycles for site preparation, which would create a 10-hour runtime work unit appropriate for World Community Grid. However, while 10 cycles was more than sufficient for the L99A T4

---

<sup>14</sup> See Chapter 6 for a full description.

lysozyme system, we performed the MF-FEB calculations on other systems that might require more cycles to reach equilibrium.

Figure 3-2 – The change in the number for waters in the site system over 25 GCMC cycles at different  $\mu$  values ( $m = \mu$ ). Each GCMC checkpoint is 1/10 of a GCMC cycle. nMax corresponds to the maximum water that could be added.

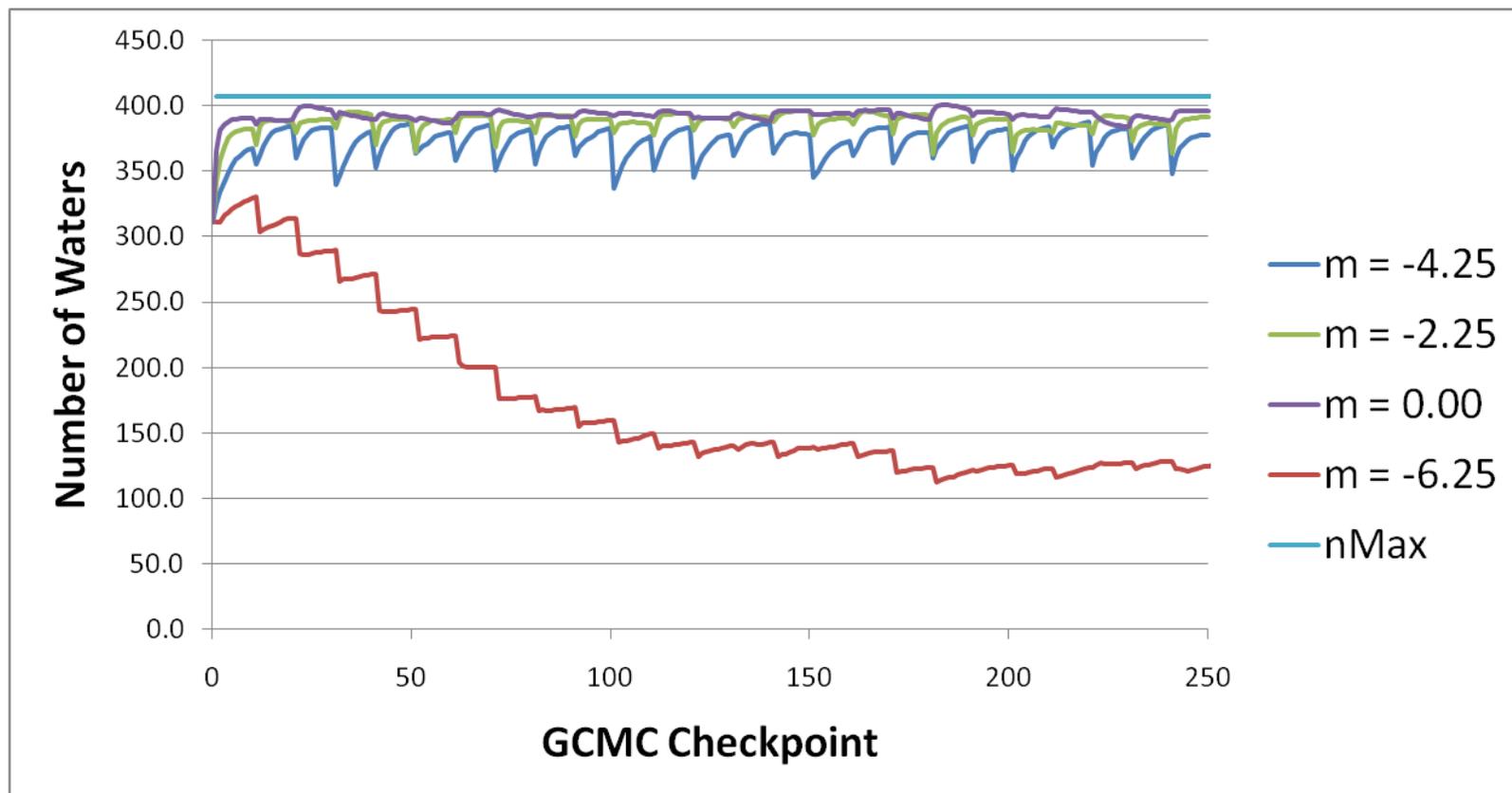
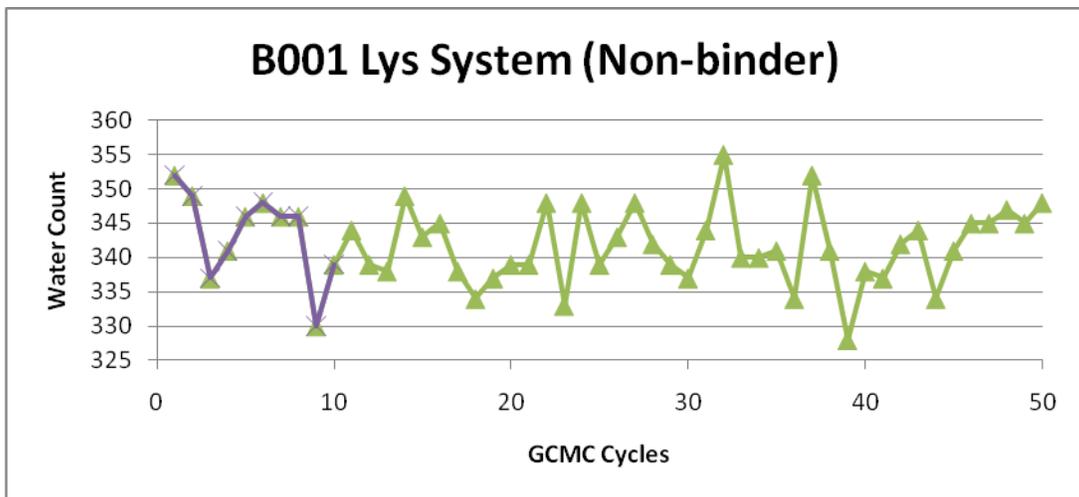
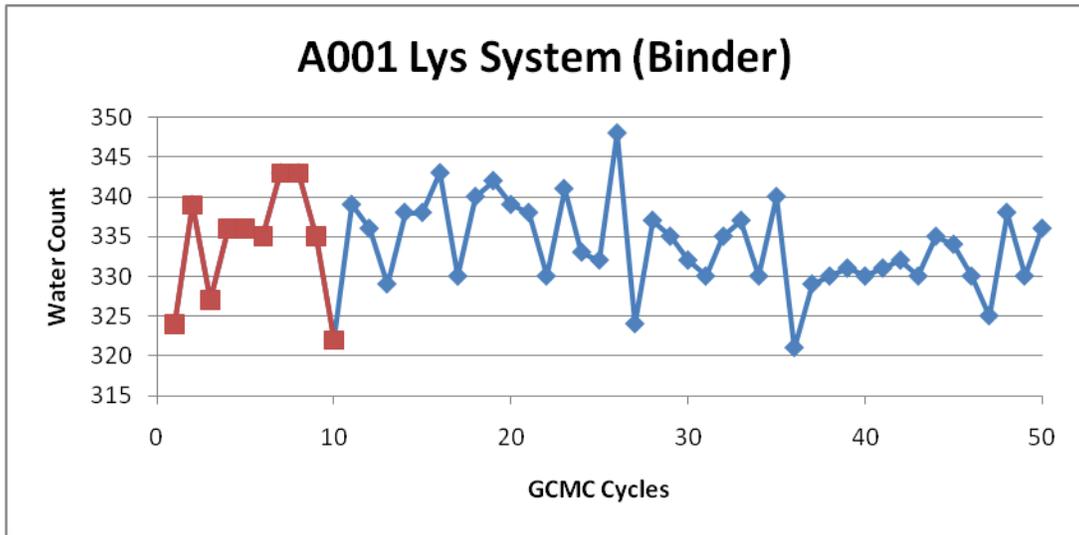


Figure 3-3 – Water count as a function of GCMC cycle for two systems with the first 10 cycles highlighted.



## **Chapter 4 Using Free energy of Binding Calculations to Improve the Accuracy of Virtual Screening Prediction**

Our hypothesis was that virtual screening could be improved by rescoring the docking generated poses using SM based FEB calculations. Our rescoring method used the well-established AutoDock4 (Huey, Morris et al. 2007; Morris, Huey et al. 2009) program to initially compute poses and empirical force field based docking scores; and Roux' mean-field pathway decoupling FEB (MF-FEB) approach (Deng and Roux 2006; Wang, Deng et al. 2006; Deng and Roux 2008) to subsequently calculate accurate binding free energies. With virtual screening and MF-FEB methods established and optimized, this chapter contains a study comparing the accuracy of virtual screening based on FEB rescoring and empirical docking scores using a test set comprised of known binders and non-binders to the engineered hydrophobic binding pocket in the L99A T4 lysozyme. (Morton, Baase et al. 1995; Morton and Matthews 1995) The L99A mutation in T4 lysozyme created a  $\sim 150 \text{ \AA}^3$  void in center of a helix bundle that bound small hydrophobic molecules. (Eriksson, Baase et al. 1992) The study asked two questions: first, could docking poses be accurately rescored using the MF-FEB calculations, and, second, could MF-FEB rescoring discriminate between binders and non-binders? Answering these two questions provided the proof-of-principle required to implement MF-FEB rescoring on a large scale.

This chapter is divided into four sections. Section 4.1 describes the methods common to all docking simulations and MF-FEB calculations presented in this chapter. Section 4.2 answers the question of using docking generated poses for MF-FEB calculations. Section 4.3 answers the question of MF-FEB rescoring's ability to discriminate between binders and non-binders. Section 4.4 reviews and discusses the findings of this chapter.

### **4.1 GENERAL METHODS**

This section describes the common methods used for calculations in this chapter.

#### **4.1.1 Hardware**

All calculations were performed at the Texas Advanced Computing Center (TACC; Austin, TX) using the Lonestar high-performance cluster. (See Section 2.1.2)

#### **4.1.2 Atomic Coordinates**

Atomic coordinates for the L99A T4 lysozyme apo-enzyme and co-crystal structures were obtained from the Protein Data Bank. The reference codes for the lysozyme co-crystal structures were: 2B6Y (apo-enzyme), 1NHB (ethylbenzene ligand), 188L (o-xylene ligand), 187L (p-xylene ligand), 182L (benzofuran ligand), 185L (indole ligand), 186L (n-butylbenzene ligand), and 184L (isobutylbenzene ligand). (Morton and Matthews 1995; Collins, Hummer et al. 2005) The benzene-lysozyme co-crystal coordinates were obtained from Dr. Y. Deng (University of Chicago).

#### **4.1.3 Docking Calculations and MF-FEB Calculations**

We performed ligand and target generation, docking calculations, virtual screening, and MF-FEB calculations using the methods described in Chapter 2. In addition, AutoGrid4 parameters were used with default settings and the search areas were centered on the lysozyme binding pocket. AutoDock4 parameters were optimized for large-scale virtual screening (as described in Section 3.1.1.3), with the number of genetic algorithm runs set to 100, the maximum number of energy evaluations set to one million, and the population size set to 200; all other docking parameters were set to default values. The most likely ligand orientation and conformation (i.e., pose) from a docking simulation was determined using a pose-based cluster analysis and a 1.5 Å cutoff and the standard pose selection method. MF-FEB calculations used the first generation CHARMM scripts with sampling time and GCMC cycles consistent with Roux and coworker previous work. (Deng and Roux 2006; Wang, Deng et al. 2006; Deng and Roux 2008)

## **4.2 ACCURACY OF MF-FEB CALCULATIONS USING AUTODOCK4-PREDICTED BINDING POSES**

In the original development of the MF-FEB, Deng et. al. used co-crystal structures to test the accuracy of the MF-FEB in predicting FEB, noting that different poses can change the FEB predictions. (Deng and Roux 2006) Additionally, previous studies have shown that docking can reproduce experimentally determined binding pose, but there was error in the predictions of  $< 2 \text{ \AA}$ . (Kellenberger, Rodrigo et al. 2004; Chen, Lyne et al. 2006; Warren, Andrews et al. 2006; Cross, Thompson et al. 2009) Therefore, we asked if docking generated poses could be used for MF-FEB calculations. Section 4.2.1 describes the test set and tests used to answer the question. Section 4.2.2 discusses the results of the tests and how the results answer our question.

### **4.2.1 Test Set and MF-FEB Calculations**

We used a “control set” to examine the accuracy of MF-FEB predictions beginning with AutoDock4 generated poses. The control set consisted of eight lysozyme ligands (n-butylbenzene, i-butylbenzene, ethylbenzene, benzofuran, benzene, indole, p-xylene, and o-xylene), each with a known binding affinity and co-crystal structure. (Morton, Baase et al. 1995; Morton and Matthews 1995) Experimental binding energies for the control set ranged from -6.70 to -4.60 kcal mol<sup>-1</sup>. (Morton and Matthews 1995) We performed three series of calculations using the control set to test the accuracy of MF-FEB using AutoDock4 generated binding poses. Each series of MF-FEB calculations were performed using coordinates from increasingly difficult docking scenarios. The first series of MF-FEB calculations were completed using coordinates extracted from inhibitor-lysozyme co-crystal structures. The second series used poses generated by docking the control set ligands into the binding site of their corresponding lysozyme structures using AutoDock4, or self-docking. The third series used poses generated by docking the control set ligands into the apo lysozyme structure using AutoDock4.

## 4.2.2 Results and Discussion

To minimize errors arising from inaccurate binding orientations between a ligand and a target protein, the first series of MF-FEB calculations were completed using coordinates extracted from inhibitor-lysozyme co-crystal structures (Table 4-1). The implicit assumption was that these structures represented an ideal docking result, thus were good starting structures for MF-FEB calculations. These MF-FEB calculations ( $\Delta\Delta G_{\text{binding}}$ ) showed good agreement to experimental measurements ( $\Delta G_{\text{binding}}$ ), with a linear correlation coefficient  $r = 0.85$  and an average error of  $1.26 \text{ kcal mol}^{-1}$  (s.d. =  $0.99 \text{ kcal mol}^{-1}$ ;  $n=3$ ) between calculated and experimental free energies (Table 4-1). During the molecular dynamics equilibration phase of the MF-FEB calculations, the initial ligand positions shifted on average  $\sim 0.7 \text{ \AA}$  relative to the crystal structures (Table 4-1). The average position during equilibration was used as the reference pose for applying translation and conformation constraints to the ligand. These calculations were more accurate than MF-FEB calculations performed with co-crystal structures as the constraint reference poses. (Wang, Deng et al. 2006)

Table 4-1 – Comparison between MF-FEB calculations and experimental binding energies using inhibitor-lysozyme co-crystal structures as the starting poses for the  $\Delta\Delta G_{\text{binding}}$  calculations. MF-FEB calculations were performed in triplicate. \*Root-mean-square deviation (RMSD) reflects the displacement of heavy atom positions between final equilibrated pose and the starting structure.

| Ligands        | Experiment  | MF-FEB calculation     |   |
|----------------|---|------------------------|---|
|                | $\Delta G_{\text{binding}}$<br>( $\text{kcal mol}^{-1}$ ) | RMSD* ( $\text{\AA}$ ) | $\Delta\Delta G_{\text{binding}}$<br>( $\text{kcal mol}^{-1}$ ) |
| n-butylbenzene | $-6.7 \pm 0.02$   | 0.9                    | $-7.8 \pm 0.5$  |
| i-butylbenzene | $-6.5 \pm 0.06$   | 0.9                    | $-9.5 \pm 0.6$  |
| ethylbenzene   | $-5.8 \pm 0.07$   | 1.3                    | $-8.0 \pm 0.2$  |
| benzofuran     | $-5.5 \pm 0.03$   | 0.3                    | $-6.5 \pm 0.4$  |
| benzene        | $-5.2 \pm 0.2$  | 0.8                    | $-5.2 \pm 1.0$  |
| indole         | $-4.9 \pm 0.06$   | 0.7                    | $-3.9 \pm 0.2$  |
| p-xylene       | $-4.7 \pm 0.06$   | 0.6                    | $-5.5 \pm 0.8$  |
| o-xylene       | $-4.6 \pm 0.06$   | 0.3                    | $-5.3 \pm 0.2$  |

Previous studies had observed that conformational bias in the T4 lysozyme binding site could reduce the accuracy of FEB predictions. (Deng and Roux 2006; Mobley, Graves et al. 2007) To determine the sensitivity of MF-FEB calculations to the initial coordinates of the ligand-protein structure, ligands of the control set were docked (or “self-docked”) into the binding site of their corresponding lysozyme structures using AutoDock4. This docking program has successfully predicted co-crystal conformations for a number of systems. (Huey, Morris et al. 2007) The ligands in this case were not energy minimized prior to the docking calculations and therefore the docking simulations inherited the information implicit in the co-crystal coordinates. Since Val 111 in the T4 lysozyme binding pocket was observed to adopt ligand-dependent conformations (Morton and Matthews 1995), self-docking was expected to reduce systematic bias arising from treating the protein binding site as a fixed structure.

Table 4-2 – MF-FEB predictions performed with self-docked poses. The root-mean-square deviation (RMSD) listed in the docking column compared AutoDock4-generated poses relative to the co-crystal structures. Only non-hydrogen atoms were used to tabulate RMSD values. The MF-FEB calculations were repeated 3 times with different starting seeds. The experimental binding data is repeated for ease of reference.

| Ligands        | Experimental   | Autodock     |                                    | MF-FEB       |   |
|----------------|--|--------------|------------------------------------|--------------|---|
|                | $\Delta G_{\text{binding}}$<br>(kcal mol <sup>-1</sup> ) | rmsd*<br>(Å) | Score<br>(kcal mol <sup>-1</sup> ) | rmsd*<br>(Å) | $\Delta \Delta G_{\text{binding}}$<br>(kcal mol <sup>-1</sup> ) |
| n-butylbenzene | -6.7 ± 0.02  | 1.0          | -5.24                              | 1.2          | -11.3 ± 1.2   |
| i-butylbenzene | -6.5 ± 0.06  | 1.7          | -4.97                              | 0.4          | -7.3 ± 0.1  |
| ethylbenzene   | -5.8 ± 0.07  | 1.3          | -4.46                              | 0.4          | -5.9 ± 0.1  |
| benzofuran     | -5.5 ± 0.03  | 2.2          | -4.77                              | 2.1          | -6.2 ± 0.5  |
| benzene        | -5.2 ± 0.16  | 1.1          | -4.00                              | 0.7          | -4.4 ± 0.7  |
| indole         | -4.9 ± 0.06  | 1.5          | -5.05                              | 0.6          | -3.8 ± 0.2  |
| p-xylene       | -4.7 ± 0.06  | 1.3          | -4.54                              | 0.6          | -7.0 ± 0.6  |
| o-xylene       | -4.6 ± 0.06  | 1.1          | -4.51                              | 0.7          | -6.4 ± 0.2  |

Relative to the co-crystal structures, the AutoDock4 generated structures and the MF-FEB constraint structures had average RMSD of 1.4 Å and 0.8 Å, respectively (Table 4-2). However, the constraint reference structures produced from either AutoDock4 or co-crystal starting structures differed by ~1 Å after the equilibration phase of the MF-FEB calculations.

AutoDock4 scores for each ligand-lysozyme complex showed little correlation with experimental  $\Delta G_{\text{binding}}$  measurements ( $r = 0.52$ ), while the MF-FEB  $\Delta\Delta G_{\text{binding}}$  calculations correlated weakly with the experimental  $\Delta G_{\text{binding}}$  measurements ( $r = 0.67$ ) (Table 4-2). The average deviation between AutoDock4 scores and experimental  $\Delta G_{\text{binding}}$  measurements were  $0.86 \text{ kcal mol}^{-1}$ , whereas the average deviation between MF-FEB  $\Delta\Delta G_{\text{binding}}$  and experimental  $\Delta G_{\text{binding}}$  measurements were  $1.53 \text{ kcal mol}^{-1}$ . However, these differences in the average deviation were not statistically different (two-tailed Student T-test,  $p=0.23$ ). These results implied that AutoDock4 scores and MF-FEB calculations would likely be unable to rank-order ligands with less than 10-fold differences in equilibrium dissociation constants.

In more demanding trials, the apo lysozyme structure was used as a target for AutoDock4 and MF-FEB calculations using the control set (Table 4-3). These simulations were characteristic of approaches used for virtual screening where numerous ligand structures were tested for binding to a single protein conformation. Similar to the above experiments with self-docking starting conformations, the AutoDock4 program positioned the test ligands within the apo lysozyme structure in orientations that approximated the experimentally determined co-crystal structures (average RMSD  $\sim 1.6 \text{ \AA}$ ) (Table 4-3). The MF-FEB molecular dynamics calculations typically adjusted the ligand positions so they more closely matched the co-crystal structures (average RMSD  $\sim 1.4 \text{ \AA}$ ) (Table 4-3). The observed displacements of the docked ligands relative to the corresponding X-ray structures were largely due to the conformation of Val 111 in the apo structure, which reduced the volume of the lysozyme apo binding site relative to the observed co-crystal structures. MF-FEB calculations performed with the AutoDock4 structures as a starting conformation differed an average of  $\sim 1.44 \text{ kcal mol}^{-1}$  from the experimental  $\Delta G_{\text{binding}}$  measurements (Table 4-3). However, the set of individual MF-FEB calculations correlated poorly with experimental  $\Delta G_{\text{binding}}$  measurements ( $r = -0.16$ ).

MF-FEB calculations initiated with either AutoDock4-generated or co-crystal ligand conformations generally produced similar values; therefore docking pose could be used for MF-

FEB calculations. On average, more accurate MF-FEB calculations occurred in cases where the AutoDock4 structures most closely reproduced the available co-crystal structures. Docking against an apo protein that had binding site conformational differences relative to the co-crystal structure produced ligand structures and that were displaced from the co-crystal generally produced less accurate MF-FEB values relative to the co-crystal structures.

Table 4-3 – MF-FEB prediction using an apo lysozyme protein structure as the docking target. For the MF-FEB calculations, three replicated calculations were performed. The experimental binding data is repeated for ease of reference.

| Ligands        | Experimental   | Docking      |                                    | MF-FEB       |  |
|----------------|--|--------------|------------------------------------|--------------|--|
|                | $\Delta G_{\text{binding}}$<br>(kcal mol <sup>-1</sup> ) | RMSD*<br>(Å) | Score<br>(kcal mol <sup>-1</sup> ) | RMSD*<br>(Å) | $\Delta\Delta G_{\text{binding}}$<br>(kcal mol <sup>-1</sup> ) |
| n-butylbenzene | -6.7 ± 0.02  | 1.6          | -6.20                              | 1.6          | -3.5 ± 0.4   |
| i-butylbenzene | -6.5 ± 0.06  | 1.7          | -5.50                              | 1.7          | -3.9 ± 0.6   |
| ethylbenzene   | -5.8 ± 0.07  | 1.4          | -4.84                              | 0.9          | -7.3 ± 0.2   |
| benzofuran     | -5.5 ± 0.03  | 2.1          | -4.69                              | 2.1          | -5.5 ± 0.6   |
| benzene        | -5.2 ± 0.2   | 1.3          | -3.72                              | 0.5          | -4.5 ± 1.1   |
| indole         | -4.9 ± 0.06  | 1.3          | -5.03                              | 1.6          | -3.0 ± 0.2   |
| p-xylene       | -4.7 ± 0.06  | 1.4          | -4.37                              | 1.8          | -4.4 ± 1.0   |
| o-xylene       | -4.6 ± 0.06  | 1.8          | -4.50                              | 0.8          | -5.9 ± 2.4   |

\*relative to the co-crystal structure

#### 4.3 ENRICHMENT AND DISCRIMINATION CHARACTERISTICS OF MF-FEB BASED VIRTUAL SCREENING

As discussed above, relatively accurate binding free energies could be calculated using the bound conformations produced by docking small flexible ligands to the apo-lysozyme. The average errors for these calculations were ~1.5 kcal mol<sup>-1</sup>, which were not sufficiently accurate to produce high correlations to experimental binding energies that spanned a narrow range of values (2.1 kcal mol<sup>-1</sup>). In contrast, for virtual screening experiments it is important to discriminate between molecules that bind to the target protein and molecules that do not bind to the target protein. For virtual screening, the free energies could include both positive and

negative values and span a larger range of values than observed for the previously examined experimental binding energies. Therefore, we investigated if the accuracy of calculated free energies was sufficient to distinguish between binding and non-binding molecules. To answer the question we performed a small-scale virtual screening. Section 4.3.1 describes the compound database, or screening set, and virtual screening method used to answer our question. Section 4.3.2 discusses the results of the small-scale virtual screening.

#### **4.3.1 Test Set and MF-FEB Calculations**

We used a screening test set, to examine the ability of MF-FEB predictions to discriminate between L99A T4 lysozyme binders and non-binders in a small-scale virtual screening. The screening set was a database of 30 compounds selected from the data of thermal shift experiments that categorized the compounds as either lysozyme binders (16 compounds) or non-binders (14 compounds). (Morton and Matthews 1995) Experimental FEB for the known binders in the screening set ranged from -6.70 to -4.59 kcal mol<sup>-1</sup>. (Morton and Matthews 1995) We docked all ligands in the screening set in to the apo T4 lysozyme L99A and rescored the docking generated poses using the MF-FEB calculations.

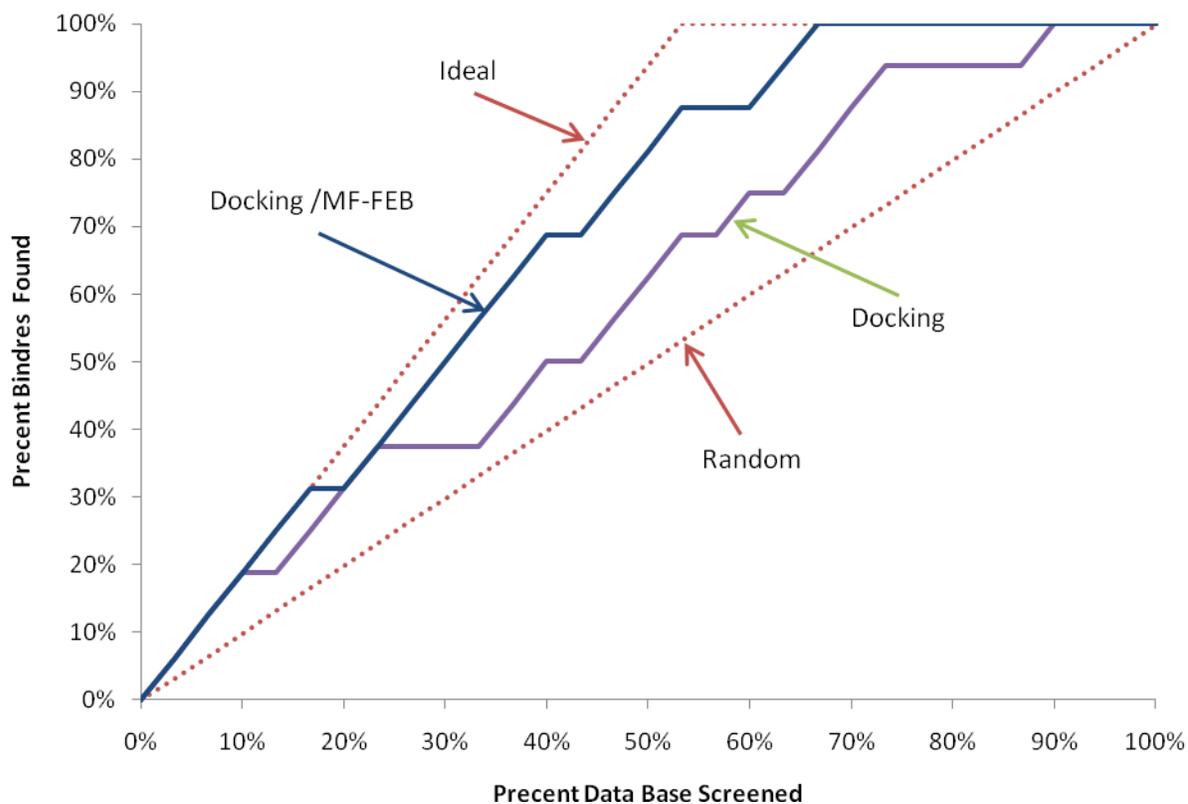
#### **4.3.2 Results and Discussion**

A small scale virtual screening (SSVS) experiment was completed using apo L99A T4 lysozyme as the target protein for screening a compound library of 16 known binders and 14 known non-binders (Table 4-4). In this experiment, a correlation coefficient could not be determined as the  $\Delta G_{\text{binding}}$  for the non-binders was unknown. The virtual screening experiment provided insight on enrichment, ligand discrimination, the responsiveness of the energy functions, and the role of solvation in binding predictions.

The information in Table 4-4 was used to construct enrichment curves that compared the percent of binders relative to the energy-ranked virtual database (Figure 4-1). Ranking the compound library by either AutoDock4 score or MF-FEB energy identified known lysozyme binders as the first two hits (Table 4-4, Figure 4-1). Moreover, the top scoring hit in each

method corresponded to one of the strongest lysozyme binders in the test set. However, after the initial three successful hits, the percentage of binders predicted by AutoDock4 scores began to approach a random distribution. In contrast, there were very few false positive hits using MF-FEB energies, and the percentage of binders identified relative to percentage of hits examined closely followed an ideal distribution.

Figure 4-1 – Enrichment curve showing docking-MFFEB enrichment compared to docking only enrichment in a virtual screening on a 30 member compound base with 16 known binders and 14 known non-binders. The ideal line represent the best case result where no false positives. The random line corresponds with a random distribution of binders and non-binders.



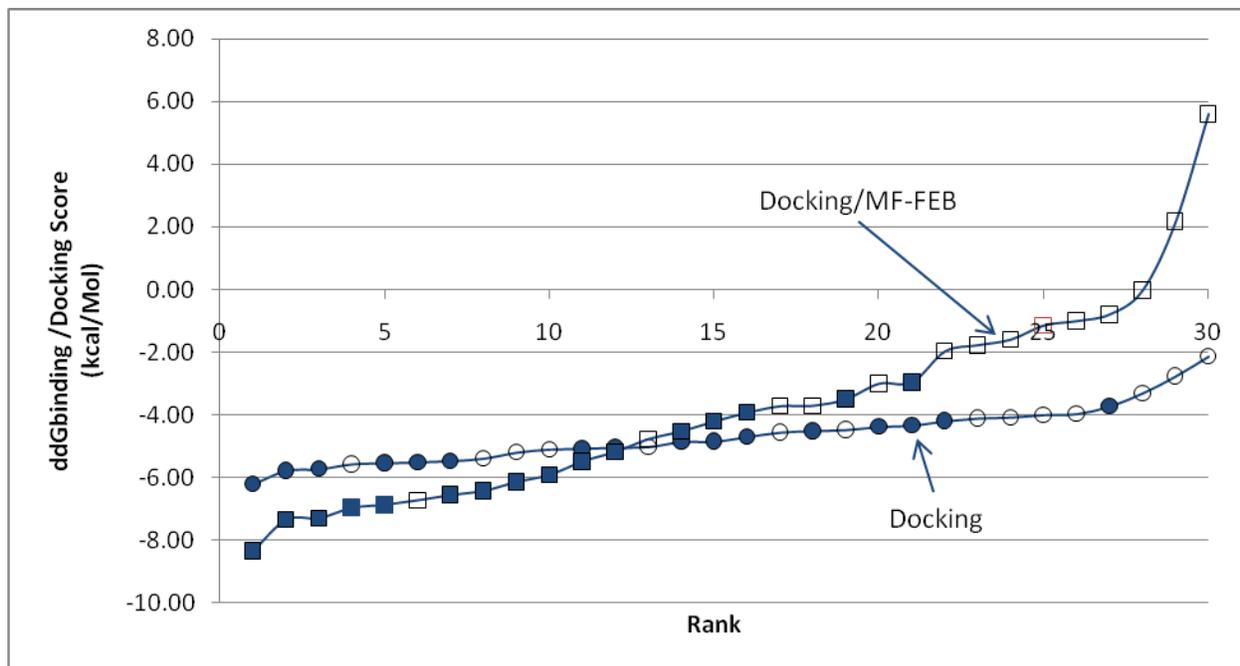
Both scoring methods predicted tert-butylbenzene to be a lysozyme binder; AutoDock4 placed this ligand as its third best binder and MF-FEB placed it as the fourth best binder. Tert-butylbenzene was an interesting non-binding ligand since it was chemically similar to the two

strongest binders in the lysozyme test-set. It is possible that tert-butylbenzene is a non-binder because it cannot enter the lysozyme binding site as opposed to having unfavorable interactions with the binding site. This restriction on binding would not be recognized by the docking or MF-FEB calculations.

Overall, the library of lysozyme binders and non-binders was more highly enriched when ordered by MF-FEB energy instead of docking score. It is difficult to compare the enrichment observed in this MF-FEB study to previous comparative enrichment studies of docking programs (Bursulaya, Totrov et al. 2003; Cummings, DesJarlais et al. 2005; Kontoyianni, Sokol et al. 2005; Warren, Andrews et al. 2006) because the number of compounds in the lysozyme inhibitor test library was small and contained similar numbers of binders and non-binders.

The  $\Delta\Delta G_{\text{binding}}$  for ligands in the co-crystal test set spanned  $2.1 \text{ kcal mol}^{-1}$ . In the virtual screening experiment, the effective range of free energy values was greater since both non-binders and binders were included. The docking calculations displayed similar distributions, with docking scores for the library compounds spanning  $\sim 4 \text{ kcal mol}^{-1}$  (Table 4-4, Figure 4-2), and docking scores for the bound ligands spanning  $2.5 \text{ kcal mol}^{-1}$  (Table 4-4, Figure 4-2). The small response range of the docking scores was consistent with previous virtual screening experiences in our laboratory (data not shown). In contrast, MF-FEB values calculated for the compound library spanned  $\sim 14 \text{ kcal mol}^{-1}$ , and the majority of non-binders were predicted to have “binding” energies less than  $-2 \text{ kcal mol}^{-1}$  (Table 4-4, Figure 4-2). MF-FEB energies were better than docking scores at discriminating between binders and non-binders. The average docking score for lysozyme binders and non-binders was  $-5.0 \pm 0.3$  and  $-4.3 \pm 0.5 \text{ kcal mol}^{-1}$ , respectively. The p-score between these two groups was 0.015, which suggested a weak statistically-relevant difference between the two groups. The average MF-FEB value for lysozyme binders and non-binders was  $-5.7 \pm 0.7$  and  $-1.6 \pm 1.6 \text{ kcal mol}^{-1}$  respectively. The p-score between these groups was 0.00009, indicating the discrimination between the two groups based on MF-FEB energies was clearly significant.

Figure 4-2 – Energy distribution of ligands rank ordered by docking score (circles) or MF-FEB energy (squares). Filled and unfilled markers represent lysozyme binders and non-binders, respectively.



The high degree of discrimination observed with the MF-FEB calculations may arise from an accurate treatment of solvation effects. The majority (11 of 14) of non-binder compounds in the test library were weakly polar, whereas the T4 lysozyme L99A binding site was predominantly hydrophobic. The free energy of binding ( $\Delta\Delta G_{\text{binding}}$ ) was the differential energy for the ligand interacting with the protein ( $\Delta G_{\text{site}}$ ) and the solvent ( $\Delta G_{\text{solvation}}$ ) (Table 4). The average  $\Delta G_{\text{site}}$  for binders and non-binders was  $-7.6 \pm 0.9$  and  $-6.2 \pm 1.8$  kcal mol<sup>-1</sup>, and the p-score between these groups was 0.096. The average  $\Delta G_{\text{solvation}}$  for binders and non-binders was  $-1.9 \pm 0.8$  and  $-4.6 \pm 1.6$  kcal mol<sup>-1</sup>, with a p-score of 0.005 between these groups. Thus, it appeared that the discrimination between binders and non-binders observed with the MF-FEB calculations was largely driven by solvation calculations, which was consistent with the physical nature of ligands and binding site. The test library contained three non-binders that lacked polar groups. Tert-butylbenzene, already discussed, ranked high (i.e., lower docking score or MF-FEB energy

relative to other ligands) in both docking and MF-FEB calculations. Cyclohexane and azulene non-binders had unfavorable MF-FEB energies due to steric constraints that resulted in weak  $\Delta G_{\text{site}}$  contributions. The test library highlighted the importance of determining both solvation and interaction energy contributions for accurate calculations of the binding free energy. In response to our opening question, rescoring docking generated poses with MF-FEB can distinguish between binders and non-binders in the L99A T4 lysozyme system.

Table 4-4 – Small-scale virtual screening results for all 30 ligand comparing experimental FEB to docking score and MF-FEB. NB= Non-binder (experimentally determined)

| Ligand               | Experimental   | Docking                            | MF-FEB   |   |  |
|----------------------|--|------------------------------------|--|---|--|
|                      | $\Delta G_{\text{binding}}$<br>(kcal mol <sup>-1</sup> ) | Score<br>(kcal mol <sup>-1</sup> ) | $\Delta\Delta G_{\text{binding}}$<br>(kcal mol <sup>-1</sup> ) | $\Delta G_{\text{site}}$<br>(kcal mol <sup>-1</sup> ) | $\Delta G_{\text{solvation}}$<br>(kcal mol <sup>-1</sup> ) |
| n-butylbenzene       | -6.70 ± 0.02   | -6.20                              | -3.5 ± 0.4   | -4.6 ± 0.2  | -1.0 ± 0.5   |
| propylbenzene        | -6.55 ± 0.02   | -5.46                              | -8.3 ± 0.6   | -9.1 ± 0.7  | -0.8 ± 0.1   |
| isobutylbenzene      | -6.51 ± 0.06   | -5.50                              | -3.9 ± 0.6   | -5.3 ± 0.5  | -1.4 ± 0.3   |
| ethylbenzene         | -5.76 ± 0.07   | -4.84                              | -7.3 ± 0.2   | -8.5 ± 0.3  | -1.2 ± 0.2   |
| thianaphthene        | -5.71 ± 0.05   | -5.07                              | -7.0 ± 0.5   | -10.0 ± 0.2   | -3.0 ± 0.4   |
| toluene              | -5.52 ± 0.04   | -4.19                              | -6.4 ± 0.4   | -7.6 ± 0.06   | -1.1 ± 0.4   |
| benzofuran           | -5.46 ± 0.03   | -4.69                              | -5.5 ± 0.6   | -9.5 ± 0.5  | -4.0 ± 0.2   |
| 4-ethyltoluene       | -5.42 ± 0.01   | -5.77                              | -7.3 ± 0.1   | -8.6 ± 0.06   | -1.2 ± 0.03  |
| benzene              | -5.19 ± 0.16   | -3.72                              | -4.5 ± 1.1   | -5.2 ± 1.2  | -0.7 ± 0.2   |
| indene               | -5.13 ± 0.01   | -5.03                              | -6.1 ± 1.0   | -9.3 ± 0.3  | -3.2 ± 0.6   |
| 3-ethyltoluene       | -5.12 ± 0.02   | -5.72                              | -6.6 ± 0.4   | -7.7 ± 0.2  | -1.1 ± 0.3   |
| indole               | -4.89 ± 0.06   | -4.85                              | -3.0 ± 0.2   | -10.2 ± 0.2   | -7.2 ± 0.3   |
| m-xylene             | -4.75 ± 0.15   | -4.33                              | -5.2 ± 0.4   | -6.1 ± 0.2  | -0.9 ± 0.3   |
| p-xylene             | -4.67 ± 0.06   | -4.37                              | -4.2 ± 1.0   | -5.3 ± 1.2  | -1.1 ± 0.4   |
| o-xylene             | -4.60 ± 0.06   | -4.50                              | -5.9 ± 2.4   | -6.9 ± 2.5  | -1.2 ± 0.4   |
| 2-ethyltoluene       | -4.56 ± 0.06   | -5.53                              | -6.9 ± 0.9   | -7.9 ± 0.5  | -1.0 ± 0.7   |
| pyridine             | NB   | -3.30                              | -1.0 ± 0.4   | -5.4 ± 0.4  | -4.4 ± 0.3   |
| phenol               | NB   | -4.10                              | -1.1 ± 0.4   | -7.4 ± 0.3  | -6.2 ± 0.2   |
| cyclohexane          | NB   | -4.07                              | -1.6 ± 1.1   | -0.2 ± 1.6  | 1.4 ± 0.5  |
| p-cresol             | NB   | -4.55                              | -2.0 ± 0.3   | -7.9 ± 0.6  | -5.9 ± 0.3   |
| 1,1-diethylurea      | NB   | -3.96                              | 2.2 ± 0.5  | -8.8 ± 0.4  | -11.0 ± 0.8  |
| aniline              | NB   | -3.99                              | -3.0 ± 0.8   | -8.7 ± 0.5  | -5.7 ± 0.3   |
| benzyl alcohol       | NB   | -4.47                              | -0.02 ± 1.3  | -7.4 ± 0.9  | -7.4 ± 0.5   |
| trans-cinnamaldehyde | NB   | -5.56                              | -1.8 ± 0.4   | -8.0 ± 0.12   | -6.3 ± 0.4   |
| tert-butylbenzene    | NB   | -5.09                              | -6.7 ± 1.3   | -8.1 ± 1.2  | -1.4 ± 0.3   |
| ethanol              | NB   | -2.13                              | 5.6 ± 0.1  | 2.3 ± 0.2   | -3.4 ± 0.3   |
| quinoline            | NB   | -5.01                              | -3.7 ± 0.5   | -10.2 ± 0.7   | -6.5 ± 0.2   |
| 1-heptoanol          | NB   | -5.19                              | -4.8 ± 0.6   | -8.0 ± 0.9  | -3.3 ± 0.6   |
| azulene              | NB   | -5.38                              | -0.8 ± 0.3   | -4.6 ± 0.3  | -3.8 ± 0.1   |
| furan                | NB   | -2.76                              | -3.7 ± 0.3   | -4.2 ± 0.2  | -0.4 ± 0.1   |

#### 4.4 CONCLUSIONS

MF-FEB energies correlated strongly with experimental binding energies when accurate structures of the lysozyme-ligand complex were used to initiate the MF-FEB calculations. As the deviation between the crystallographic and docked structures increased, the average error of the MF-FEB calculations increased. However, small-scale virtual screening results showed that rank-ordering libraries based on MF-FEB energies of docking poses could provide significant enrichment relative to rank-ordering based on docking scores. The MF-FEB rescoring provided significant discrimination between micromolar binders and non-binders. These results suggest that the false positive rate of conventional virtual screening projects could be significantly reduced by coupling docking poses with MF-FEB calculations.

Despite the promising discrimination observed with the lysozyme small scale virtual screen, the MF-FEB calculations suffered from many of the drawbacks common to all molecular calculations. The available force fields were approximate, and could be improved by including polarizable force fields. (Mobley, Chodera et al. 2006; Deng and Roux 2009) Convergence of the molecular dynamics was a concern, since the variations in replicate MF-FEB calculations suggested that some systems were not fully converged. However, in spite of these limitations, the use of MF-FEB energies with preliminary AutoDock4-determined poses provided a powerful approach to discriminate between binders and non-binders.

In addition, the sensitivity of these calculations to the initial docked structures was determined. Receptor binding site flexibility is a challenge for docking programs. (Teodoro and Kavraki 2003; Kitchen, Decornez et al. 2004; Ghosh, Nie et al. 2006; Leach, Shoichet et al. 2006; Totrov and Abagyan 2008) Some approaches have modeled receptor flexibility by combining docking calculations with receptor molecular dynamics (MD). (Alonso, Bliznyuk et al. 2006) However, the MD performed during the above MF-FEB calculations were not sufficiently lengthy to simulate changes in the receptor binding site conformation. More

accurate FEB predictions may result from calculations that incorporate receptor flexibility. (Mobley, Chodera et al. 2007; Mobley, Graves et al. 2007) Significantly, MF-FEB calculations initiated with either crystal structures or AutoDock4 poses were in good agreement with experimental measurements. In this limited virtual screening experiment, false-positive and enrichment rates are improved when compounds are evaluated with FEB rescoring as opposed to AutoDock4 scores. Moreover, rescoring allows micromolar binders to be separated from non-binders. These results suggest that MF-FEB rescoring can be used to efficiently screen large virtual libraries for novel drug leads.

Currently, the limited accuracy of MF-FEB calculations initiated from docking poses may prevent these calculations from being used to design ligands with incremental improvements (e.g., 1.5 kcal mol<sup>-1</sup>) in binding energies. However, these calculations could discriminate between binders and non-binders, and thus provided significant enrichment of libraries. The results of this study were encouraging and merit further testing against larger virtual libraries.

## **Chapter 5 Discovering Dengue Drugs-Together and Influenza Antiviral Drug Search**

Chapter 4 established that we could use MF-FEB calculations to rescore docking generated poses and that rescoring increased enrichment in a small test set. However, to more fully test MF-FEB rescoring required the use of large drug-like test sets. Large scale testing required significant computational power. Both docking and MF-FEB calculations are amenable to grid computing as discussed in Section 1.6. Therefore, we launched two projects on IBM's World Community Grid: Discovering Dengue Drug-Together (DDDT), and Influenza Antiviral Drug Search (IADS). These projects not only provided the computational power to test the MF-FEB rescoring, but also allowed to us to utilize virtual screening with MF-FEB rescoring in an important drug discovery effort.

This chapter describes DDDT and IADS. Section 5.1 provides an overview of the process of virtual screening with MF-FEB rescoring. Section 5.2 describes IBM's World Community Grid. Section 5.3 discusses DDDT. Section 5.4 discusses IADS.

### **5.1 VIRTUAL SCREENING WITH AUTODOCK4 AND MF-FEB RESCORING**

As discussed in Section 1.6, the central hypotheses of this work is that rescoring docking generated poses using MF-FEB calculations will improve enrichment over docking scoring functions alone. To test and apply our hypotheses, we launched the DDDT and IADS drug discovery projects on World Community Grid. Figure 5-1 shows the workflow used for the two drug discovery projects. The drug discovery projects began with virtual screening using AutoDock4 followed by rescoring using MF-FEB calculations. For convenience of implementation, we divided each project into two phases. Phase 1 was the virtual screening using AutoDock4. Phase 2 was the rescoring using MF-FEB calculations. Sections 5.1.1 and 5.1.2 provide an overview of Phase 1 and Phase 2 respectively. Chapter 6 provides a detailed explanation of the methods used in Phase 1 and Phase 2.

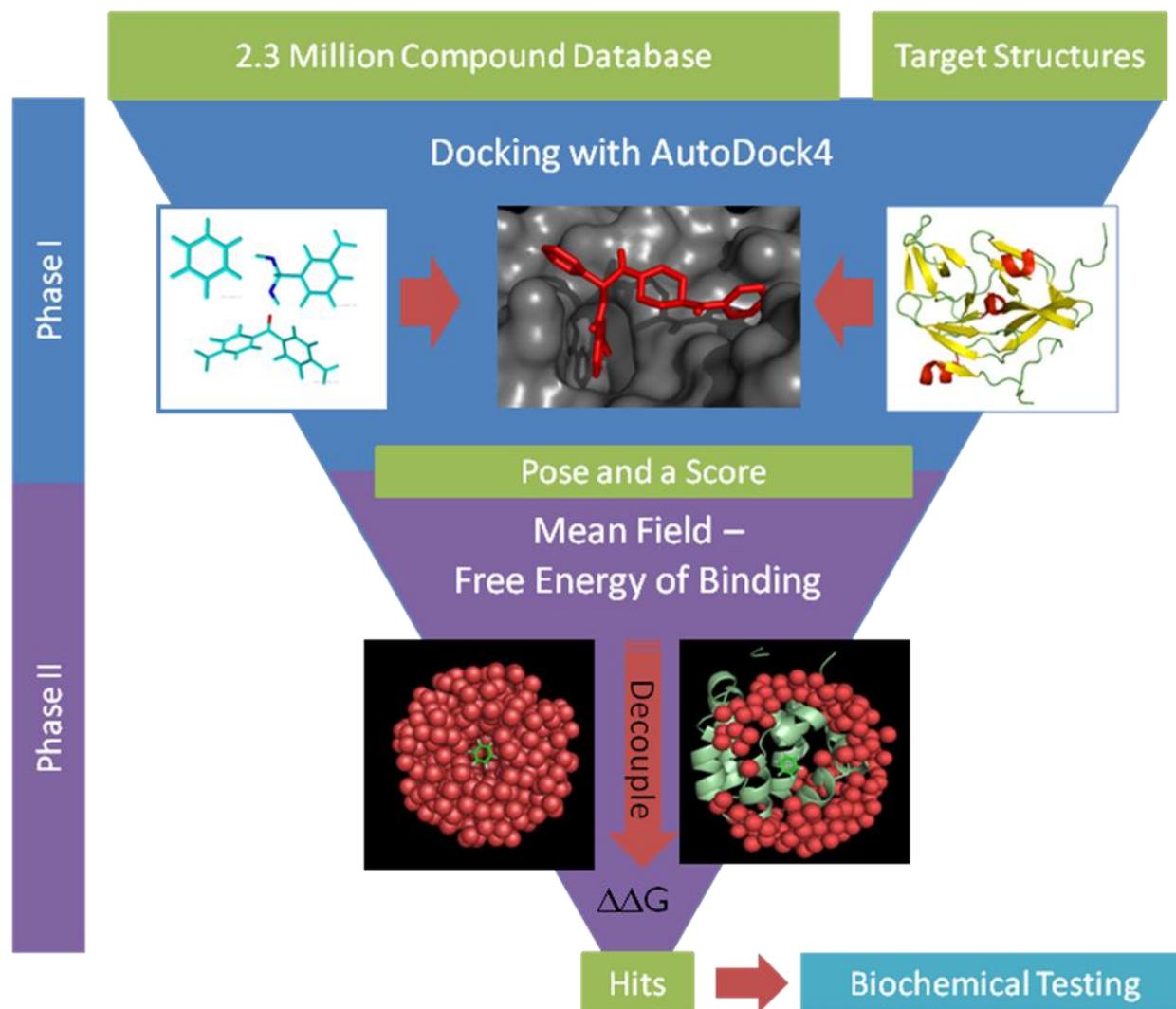
### **5.1.1 Phase 1**

Phase 1 was the virtual screening phase, which generated poses and initial compound ranks using AutoDock4. The compound database used for the virtual screening was obtained from the ZINC database (Irwin and Shoichet 2005) and contained ~2.3 million compounds. The virtual screening targets were the protein drug targets selected for each drug discovery project. Each compound in the database was docked to each target using AutoDock4 (Morris, Goodsell et al. 1998; Huey, Morris et al. 2007) using the docking parameters determined in Section 3.1.1.3. We selected a single result pose for each docking simulation using the standard selection method described in Section 3.1.1.2. The docking results were filtered as described in Section 3.1.2 to obtain the top 2,000 compound by score and the top 2,000 compounds by cluster size. Additionally, we selected a small portion of the top ranked compounds from Phase 1 for early testing.

### **5.1.2 Phase 2**

Phase 2 was the rescoring of the 4,000 selected docking poses from Phase 1 using MF-FEB calculations developed by Roux et al. (Woo and Roux 2005; Deng and Roux 2006; Deng and Roux 2008) While World Community Grid provided significant computational resources, it did not provide sufficient resources to rescore all 2.3 million compounds, thus we could only rescore a small fraction of the total virtual screening results. A selected ligand and target from Phase 1 were first processed into inputs usable by CHARMM for the MF-FEB calculations. We then performed the MF-FEB calculations using the parameters determined in Section 3.2 and re-ranked the ligands based on the computed FEB. The top ranking compounds were then tested for activity.

Figure 5-1 – Work flow diagram for virtual screening with AutoDock4 and MF-FEB rescoring.



## 5.2 WORLD COMMUNITY GRID

The World Community Grid’s mission is to “create the largest public computing grid to tackle projects that benefit humanity”.

([http://www.worldcommunitygrid.org/about\\_us/viewAboutUs.do](http://www.worldcommunitygrid.org/about_us/viewAboutUs.do)) It is philanthropic effort of IMB that provides the resources and talent to build and maintain a global grid to aid researchers in porting their projects to the grid. The grid itself is composed of over half a million members,

who donate processor time on over 1.7 million devices. (<http://www.worldcommunitygrid.org/stat/viewProjects.do>) Each member device on the grid uses the BOINC (<http://boinc.berkeley.edu/>) client to receive and run work. To date, members have donated over 466,000 years worth of CPU time to the projects of World Community Grid. (<http://www.worldcommunitygrid.org/stat/viewProjects.do>)

Seventeen projects have run or are currently running on World Community Grid, including some of the following examples. The Help Cure Muscular Dystrophy project uses protein-protein docking methods to identify potential drug targets among the more than 200 genes involved in neuromuscular disease. (<http://www.worldcommunitygrid.org/research/hcmd/overview.do>) The Nutritious Rice for the World project used protein-folding methods to help develop new rice strains. (<http://www.worldcommunitygrid.org/research/rice/overview.do>) The Computing for Clean Water project uses MD simulations to study water flow to develop new nanotube based water filters. (<http://www.worldcommunitygrid.org/research/c4cw/overview.do>) The FightAIDS@Home project uses AutoDock4 to find new drugs to treat HIV infections. (<http://www.worldcommunitygrid.org/research/faah/overview.do>)

The architecture of World Community Grid is composed of three tiers. The first tier is the member devices that donate CPU cycles to perform calculations for the projects using the grid. The second tier is World Community Grid's servers that organizes and submits jobs to the member devices from work units generated by the third tier, and validates the results of the work units returned from the member devices and returning them to the third tier. The third tier is the project servers that prepare work units for the grid, and evaluate the results of the calculations.

We used World Community Grid for our DDDT and IADS projects. We selected World Community Grid not only because it could provide the computational resources need for virtual screening with MF-FEB rescoring, but because they also provide the expertise to implement grid

based projects. While we implemented the projects on the project servers and were responsible for the analysis of the results, the experts at World Community Grid ported CHARMM and AutoDock4 to different operating systems, managed the workflow of submitting jobs to the member devices, and recruited new members to maintain and grow the grid. In short, we did the drug discovery, and IMB ran the grid.

### **5.3 DISCOVERING DENGUE DRUGS-TOGETHER**

This section describes our World Community Grid drug discovery project DDDT. Section 5.3.1 outlines the purpose and goals of the project. Section 5.3.2 discusses the drug discovery targets selected for the project. Section 5.3.3 concludes the section with the status of the project.

#### **5.3.1 Project Goals**

Dengue is a mosquito-borne virus of the family *Flaviviridae* that infects humans causing dengue fever, dengue hemorrhagic fever and potentially fatal dengue shock syndrome. (Tomlinson, Malmstrom et al. 2009) With about 2.5 billion people at risk of infection and 1.5 million infected each year, there is currently no vaccine or antiviral treatment for dengue infections. (Tomlinson, Malmstrom et al. 2009) While, viewed mainly as a tropical disease, dengue virus has begun to threaten the United States. (Morens and Fauci 2008) The goal of DDDT was to identify compounds to develop into drugs to treat dengue viral infections, and the related viral infection of West Nile virus and HCV. Based on our previous success with virtual screening to identify potential antiviral for dengue (Tomlinson, Malmstrom et al. 2009), we utilized the computational power of World Community Grid to perform virtual screening with MF-FEB rescoring against three viral protease targets and one control target.<sup>15</sup>

---

<sup>15</sup> We virtually screened an additional nine Leishmania targets as a pilot study for Dr. Muskus at the University of Columbia, Medellin, under DDDT, only performing the Phase 1

### 5.3.2 Virtual Screening Targets

Flaviviruses are single-stranded positive-sense RNA viruses whose genome encodes a single poly-protein containing both structural proteins that form the virion, and non-structural proteins that are involved in the maturation of the virion and replication of the genome. (Tomlinson, Malmstrom et al. 2009; Noble, Chen et al. 2010) The ribosome translates viral poly-protein into the ER membrane with different sections of the protein being on opposite sides. (Tomlinson, Malmstrom et al. 2009; Noble, Chen et al. 2010) Cellular and viral proteases then process the poly-protein into its separate protein parts. The viral serine protease is at the n-terminus of the NS3 (Non-Structure) protein and cuts a basic residue, like trypsin, and requires a cofactor for activity, the NS2B for West Nile and dengue, and the NS4A for HCV. (Yusof, Clum et al. 2000; Erbel, Schiering et al. 2006; Tomlinson, Malmstrom et al. 2009) The viral protease is required for viral replication and therefore is a potential drug target. (Tomlinson, Malmstrom et al. 2009; Noble, Chen et al. 2010) Because crystal structures have been solved for the all three protease it is also an attractive target for SBDD. Inhibitors have been identified for the West Nile protease (Tomlinson and Watowich 2008), dengue protease (Tomlinson, Malmstrom et al. 2009; Frecer and Miertus 2010; Tomlinson and Watowich 2011), and extensive drug discovery and development work has been done to find protease inhibitors for HCV protease (Naggie, Patel et al. 2010). We therefore selected the NS3 protease as our drug discovery target.

We targeted the volume around the catalytic, P1, and P2 sites of the NS3 proteases for docking simulations. For the dengue protease, we used the 2FOM structure of the apo-NS2B/NS3 protease from dengue 2 and the 2VCB full NS3 from dengue 4. For the West Nile protease, we used the 2FP7 and 2IJP structures of an inhibitor bound NS2B/NS3 protease. For HCV protease, we used the 2A4R structure of the NS3 protease and NS4A cofactor. We also included the 1EB2 trypsin structure to identify any promiscuous ligands.

---

AutoDock4 virtual screenings. All virtual screening were successful and the results given to Dr. Muskus for analysis.

### **5.3.3 Current Progress Report**

We have completed Phase 1 of DDDT taking 11,737 years of CPU time. We are currently running Phase 2, and are rescoring the results from the 2FOM and 2FP7 structures following the completion of the testing stage. Phase 2 has so far used over 1,200 years of CPU time.

## **5.4 INFLUENZA ANTIVIRAL DRUG SEARCH**

This section describes our World Community Grid drug discovery project IADS. Section 5.4.1 outlines the purpose and goals of the project and discusses the drug discovery targets selected for the project. Section 5.4.2 concludes the section with the status of the project.

### **5.4.1 Project Goals**

In response to the H1N1 influenza pandemic of 2009, and at the request of World Community Grid, we launched the IADS. We selected the neuraminidase as our drug discovery target. Neuraminidase removes sialic acid from the glycoproteins and glycolipids on the cell membrane to release the virion from the cell and is the target for drugs zanamivir and oseltamivir. (Gamblin and Skehel 2010) We targeted the sialic acid binding site of neuraminidase for docking simulations. We used the following neuraminidase structures for virtual screening: 2HU4, 2HTY, 3B7E, 2QWE, 1L7G, 3CKZ, 3CLZ, 2HTQ, and 2HTW.

### **5.4.2 Current Progress Report**

We have completed Phase 1 of IADS, taking 2,876 years of CPU time. While, a Phase 2 is planned for IADS it is on hold until the completion of the first set of Phase 2 results for DDDT. No compounds have been tested from Phase 1.

## **Chapter 6 Grid Implementation of Virtual Screening and Mean Field-Free Energy of Binding Calculations**

Chapter 5 described how large-scale virtual screening projects using MF-FEB rescoring could be performed using IBM's World Community Grid. This chapter describes the methods used for DDDT and IADS projects. Section 6.1 describes the method used for Phase 1 virtual screening. Section 6.2 describes the methods used in Phase 2 MF-FEB rescoring.

### **6.1 PHASE 1 – VIRTUAL SCREENING**

The first phase of our World Community Grid projects was the virtual screening of the targets described in Sections 5.3.2 and 5.4.1 against a compound database of 2.3 million compounds. This section describes the methods used to perform Phase 1 virtual screenings. Section 6.1.1 describes the methods used to prepare the compound database and targets for virtual screening on World Community Grid. Section 6.1.2 describes how we ran virtual screening projects on World Community Grid. Section 6.1.3 compares the runtime for virtual screening project on World Community Grid versus using supercomputers. Section 6.1.4 describes how we processed and stored the virtual screening results for Phase 2.

#### **6.1.1 Preparing for Phase 1 Virtual Screening**

To perform a virtual screening project on World Community Grid, we first prepared the compound database, the targets, and the parameter files for AutoDock4 and AutoGrid4. Section 6.1.1.1 describes the preparation of the compound database. Section 6.1.1.2 describes target preparation. Section 6.1.1.3 describes the preparation of the AutoDock4 and AutoGrid4 parameter files.

##### ***6.1.1.1 Building the Compound Database***

We used a subset of the ZINC database (2007 version) for our compound database for our World Community Grid projects. We used three filters to parse the ZINC database. The first filter was “reputable dealers.” The ZINC database was built from the product catalogues of

numerous chemical companies worldwide. The “reputable dealers” filter selected compounds from chemical companies that, according to the ZINC database, provide quality compounds (i.e., pure and correct). We selected this filter to increase the likelihood that a compound selected for testing could be purchased and would be of a reasonable quality for reliable testing.<sup>16</sup> The ZINC database was further filtered to include only drug-like (Lipinski, Lombardo et al. 2001), lead-like (Teague, Davis et al. 1999), or both compounds, the second and third filters respectively. These two filters increased the likelihood of a compound selected for testing to have water solubility and drug-ability.

We downloaded the filtered ZINC library multi-compound .mol2 files. We processed multi-compound .mol2 files to *ligand.pdbqt* files according to the procedure outlined in Section 2.3.2.1. The final compound database contained 2,253,582 compounds, including tautomers.

In order to find compounds quickly and to build work groups, we subdivided the *ligand.pdbqt* files into groups. The *ligand.pdbqt* files were first divided into 902 groups of ~2,500 compounds per group. These groups were further divided into letter groups, “a” through “p,” with each letter group containing ~60 of the 902 groups.<sup>17</sup>

#### **6.1.1.2 Target Preparation**

We obtained the crystal structures for all targets described in Sections 5.3.2 and 5.4.1 from the Protein Data Bank. We then transposed the coordinates of each target placing the

---

<sup>16</sup> In practice, this tended not to be the case because “reliable” dealers sometimes obtained their compound from “unreliable” sources, although the sources of compounds are disclosed.

<sup>17</sup> Originally, before the work was moved to TACC, the first division was “a” to “o” based on the 15 nodes of hpcluster1. On Ranger, there were 16 processors per node, therefore “a” to “p.”

origin at the active site or the geometric center of a co-crystal ligand. The targets were then prepared as previously described in Sections 2.2.2, 2.3.1.1.2, and 2.3.2.2.

### **6.1.1.3 AutoGrid4 and AutoDock4 Parameters**

We generated template temp.gpf files for AutoGrid4 for each target based on the search areas described in Sections 5.3.2 and 5.4.1 according to the standards described in Section 2.3.1.2. We generated template temp.dpf files according to the virtual screening parameters established in Section 3.1.1.3.

## **6.1.2 Virtual Screening on World Community Grid**

The process of virtual screening on World Community Grid requires that prepared docking simulations be submitted to World Community Grid servers. World Community Grid servers sent the docking simulations out to member devices for processing, then collected the results, and returned the results to the project servers. We then processed the returned results for use in Phase 2. This section describes the methods used to complete the three steps above. Section 6.1.2.1 describes the preparation of work on the project servers. Section 6.1.2.2 describes how World Community Grid performed the docking simulations. Section 6.1.2.3 describes the preparation of the virtual screening results for Phase 2.

### **6.1.2.1 Preparing Work for and Submitting Work to World Community Grid**

We submitted the docking simulations to World Community Grid servers as GNU-zipped tape archive records (i.e. .tar.gz), which contained 2,500 simulations, corresponding to one of the 902 subgroups of the compound database, as described in Section 6.1.1.1. Each collection of simulations was called a work group. Work groups were prepared from the *ligand.pdbqt* files in the compound database and the *target.pdbqt*, *AD4parameter.dat*, *temp.gpf*, and *temp.dpf* files stored in the target directory. The scripts *cf090105.py* and *cf-090503\_flu.py*, corresponding to DDDT and IADS projects respectively, processed the input files into work groups. These two scripts were the progenitor scripts to *cf-100128.py* and functioned as previously described in Section 2.3.2.2 with the following differences: (1) the scripts prepared work groups based on the

letter division of the compound database (i.e., the “a” work groups were all made at once), and (2) the output did not include shell scripts to launch AutoGrid4 and AutoDock4. The scripts outputted two files for each work group. The first was the *workgroup.tar.gz* containing each docking simulation in a sub directory. The second was a text file containing the md5 check sum for the *workgroup.tar.gz*.

Ranger supercomputer at TACC served as the project server preparing the work groups. We submitted jobs to build work groups on Ranger using a .sge script that allowed for the submission and execution of multiple serial jobs over a number of nodes. Completed work units were stored on TACC’s Corral storage system.

The main challenge faced in preparing work groups was that building the work units required a high performance computing environment. Using four nodes on Ranger, a complete virtual screening experiment could be prepared overnight; however, with one processor it would take about a week. Reading and writing to the disk, not the number of cores per node, determined the number of nodes required to prepare work groups. Due to Ranger’s architecture with four quad-core processors per node, only four jobs could be run per node, one per processor. Running over four jobs per node increased the runtime as each core waited for disk access since the script’s main functions were copying and compressing files.

#### **6.1.2.2 Docking on World Community Grid**

AutoGrid4 and AutoDock4 were implemented previously on World Community Grid for the Fight AIDS@Home project. (Chang, Lindstrom et al. 2007) Therefore, World Community Grid had an established system for handling work groups when we launched DDDT and IADS. The process for running the docking simulations started with World Community Grid servers downloading the prepared work groups from Corral at TACC. World Community Grid servers then repackaged the individual docking simulations into jobs that it sent out to member devices. Each job was designed to have a run time of approximately 10 hours on a member’s device.

World Community Grid servers packaged a varying number of docking simulations into a single job whose total calculated runtime, based on the complexity of the ligand, was 10 hours. The average number of ligands per job was five. The member devices ran AutoGrid4 and AutoDock4 for each ligand in the job, compiling all docking outputs into a single *result.dlg*, and returning only the *result.dlg* to World Community Grid servers. World Community Grid servers then compiled all the *results.dlg* files from the simulations of one work unit and returned the validated results to Corral at TACC.

Performing virtual screening in a grid environment provided different challenges than those encountered in a supercomputing environment. One challenge was processor error. World Community Grid is composed of a large number of member devices with different processors. While the error rates of any one processor is low, on the scale of World Community Grid it becomes a factor, as results from different processes are compared to each other. Originally, World Community Grid servers submitted two copies of each job grid, each with the same random number seed. The results of the matched jobs were accepted only if the matched jobs provide the same results. However, during the course of the project a test calculation was added to each job to validate the results, removing the need to submit matching jobs. The calculations were run before and after the docking simulations to check heat induced error in the processor. The other challenge was download and upload time between World Community Grid servers and member devices. As World Community Grid jobs were designed to run in the background on member devices, communication between World Community Grid servers and the member devices were kept to a minimum. Additionally, World Community Grid servers have a finite bandwidth. The most CPU efficient implementation of AutoGrid4 would be to run it once for a virtual screening project generating all required scoring grids for all the simulations. However, the scoring grid map files are about 1MB each, so uploading the grid files to the member computers, even compressed, would require significant bandwidth over hundreds of thousands of

jobs. Therefore, AutoGrid4 was run on the member devices and only the *result.dlg* was returned from each member devices, exchanging efficiency for a lower overall bandwidth requirement.

### **6.1.2.3 Processing Results**

Validated docking results from single work units were processed on the project server, Ranger, to extract the individual docking results and generated a table of results. The results returned from World Community Grid contained the *results.dlg* files from each job submitted to member computers. The *dlg\_reader.py*<sup>18</sup> script decompressed and parsed each *results.dlg* file into multiple *ligand.dlg* files containing the results of the single docking simulation. The script then processed the new *ligand.dlg* files using *docking\_run.py* as described in Section 2.3.2.3 generating the same output. The *dlg\_reader.py* script compiled the results of work unit into a single tab delineated ASCII text file with one docking result per line. We archived the unprocessed *results.dlg* on Random's storage system.

Because the docking simulations were run in the background on member devices, World Community Grid modified AutoDock4 to pause as need. Sometimes the pause would require the current docking run to be stored in the RAM. Other times, the pause would require AutoDock4 to shutdown and then start the docking simulation anew. As each docking simulation was composed of 100 GA runs, World Community Grid used GA runs as checkpoints at which to restart the docking simulation. For example, if the first 10 runs were complete before a shutdown, AutoDock4 would start on run 11. However, AutoDock4's post simulation cluster analysis would only process results obtained since the last shutdown; therefore, the *docking\_run.py* script had to re-cluster the docking results.

---

<sup>18</sup> The *dlg\_reader\_05.py* described in section 2.3.2.3 was a simplified version of *dlg\_reader.py*.

### **6.1.3 Virtual Screening Run Times on World Community Grid**

Because World Community Grid was composed of member devices donating their unused CPU time, running a single docking job was inefficient. The average wall-clock time to run 10 docking simulations was about 12 hours. The average time to finish a job on a member device from receiving the job to returning it to World Community Grid servers was 3 days, a six fold increase, not including the time required to prepare the work, transfer it to World Community Grid servers, receive the results from World Community Grid servers, and process the results. The main advantage of the grid was not its efficiency but the volume of work World Community Grid performed. The calculated time for a virtual screening experiment using our in-house 15-node cluster, hpcluster1, was 12 years, compared to 2 months for one virtual screening on World Community Grid. While a super computer like Ranger could match or beat the two months, it would require dedicating the whole system to one project.

The distribution of job completion times for a single work group was skewed, with ~90% of the jobs completed within three days and the remaining 10% taking up to two weeks to complete. The overall completion time could therefore be reduced by only submitting jobs to the more efficient member computers.

### **6.1.4 Results Extraction**

The results files generated from World Community Grid virtual screening were processed into two compound collections each containing 2,000 compounds corresponding to the lowest scoring compounds and the largest cluster compounds selected as described in Section 3.1.2. The sorted\_results\_v03.py script processed the result files. The script first generated a list of all 2.3 million compounds containing only the compound name, docking score, and cluster size. RAM limitations prevented loading the full results record into the list. The list was then sorted into two new lists, one ordered by score then cluster size, and the second by cluster size then score. The script removed the top compound from each ordered list then added it to the results lists skipping compounds already present in one of the results lists. The selection process

continued until it generated two lists of 2,000 compounds. The parallel list building process generated two unique rank ordered lists. The script read the results files for a second time extracting the full docking results and writing the results to a final tab delineated *results.txt*, thereby generating input for Phase 2.

## **6.2 PHASE 2 – GRID BASED MF-FEB CALCULATIONS**

This section describes and discusses the methods we used to perform MF-FEB calculations using World Community Grid for Phase 2 of DDDT and IADS. Section 6.2.1 describes the preparation of the files required for MF-FEB calculations. Section 6.2.2 describes the process and scripts we used to perform MF-FEB calculations on World Community Grid. Section 6.2.3 concludes by discussing the runtime and failure rate of MF-FEB calculations on World Community Grid.

### **6.2.1 Work Preparation**

This section describes the required input files and the preparation methods for performing MF-FEB calculations on World Community Grid. Section 6.2.1.1 describes the generation of the compound databases. Section 6.2.1.2 describes the preparation of the target structure. Lastly, Section 6.2.1.3 describes the required parameter and CHARMM files.

#### **6.2.1.1 Preparing Ligands**

The scripts used to prepare the MF-FEB calculations for World Community Grid required two matching compound databases in separate directories. One directory contained *ligand.mol2* files and the second directory contained *ligand.pdbqt* files. The *ligand.mol2* were copies of those obtained from the ZINC database as described in Section 6.1.1.1 and the *ligand.pdbqt* corresponded to the docking input files used in Phase 1 as described in Section 6.1.1.1. The ligands in both directories corresponded to the ligands selected for rescoring as described in Section 6.1.4. The job building scripts, described in Section 6.2.2, required the fully

protonated model of the ligand contained the *ligand.mol2* file, and the torsion bond definitions used by AutoDock4 for the docking simulations in Phase 1 contained the *ligand.pdbqt* file.

The *lib\_extractor.py* script generated the two ligand input directories. The script read in a *results.txt* file from Phase 1, and found the corresponding ligand files in the master compound databases used for Phase 1. The script then copied the ligand files to either the mol2 or the pdbqt directories, generating the input compound databases for the MF-FEB calculations.

### **6.2.1.2 Preparing Target**

We prepared the target files using the methods described in Section 2.4.1.2. We gathered the resulting *const\_dihe.str*, *target\_XXXX\_water.crd*, *disu.str*, and *seg.str* files into an input directory named after the World Community Grid job name.

### **6.2.1.3 MF-FEB CHARMM Scripts and Parameter Files**

The MF-FEB calculations required the following CHARMM script and input files: *const.str*, *par\_all22\_prot\_cmap.inp*, *pick.str*, *radius.str*, *restr\_sel.str*, *rstr\_def\_rms\_fluc.str*, *site\_pert.inp*, *site\_setup\_eqlb.inp*, *site\_stup\_gcmc.inp*, *site\_stup\_gsbp.inp*, *solv\_pert.inp*, *solv\_stup\_eqlb.inp*, *top\_all22\_prot\_cmap.inp*, *water\_400.crd*, *water\_8000.crd\_nocopy*, and *wham.inp*. Section 2.4.2 describes the use of each of the list files in the MF-FEB calculations. The files were stored together in the “inputs” directory.

A “*mock\_xmls*” directory contained template XML files. These XML files stored parameters for the MF-FEB calculations and lists of the individual job input files. The *template.xml* file contained all the parameters for the MF-FEB calculations including the MD simulation lengths, system size, perturbation steps, etc. The *template.xml* file is similar to the *job.xml* file described in Section 2.4.3.1, and was used to generate unique xml files for each ligand-target system. In addition to the *template.xml*, the *mock\_xmls* directory had seven template job xml files that contained information on what files were required for each stage of the MF-FEB calculations, which files were to be returned, and a place holder for the CHARMM

command line parameters. The template job xml files were: site\_pert\_cdqr.xml, site\_pert\_rmsd.xml, site\_stup\_equlb.xml, site\_stup\_gcmc.xml, site\_stup\_gsbp.xml, solv\_pert\_qqdr.xml, and solv\_stup\_eqlb.xml.

## **6.2.2 Running MF-FEB on World Community Grid**

Section 2.4.2 describes the computational method used for performing MF-FEB calculations, including the required inputs and the generated outputs using the second-generation CHARMM scripts used to perform the MF-FEB calculations on World Community Grid. Due to the scale of the MF-FEB rescoring projects and the architecture of World Community Grid, we employed modified methods of running the MF-FEB calculation as described in Section 2.4.3. This section describes the process of performing MF-FEB calculations on World Community Grid in three sections. Section 6.2.2.1 describes the modifications made to CHARMM input scripts to facilitate running them in the grid environment. Section 6.2.2.2 outlines the process for performing MF-FEB calculations on World Community Grid. Section 6.2.2.3 described the scripts used to manage jobs, create jobs, and calculate the results.

### ***6.2.2.1 Modifying CHARMM Inputs***

As discussed in Section 6.1.2.2 and 5.2, World Community Grid cycles harvests the member devices to perform calculations. Therefore, calculations needed to be paused and restarted. While pausing the MF-FEB calculations and retaining them in memory to be restarted required no modification of the CHARMM scripts, having to checkpoint and fully restart the calculations required modifying them. We modified the CHARMM scripts site\_pert.inp, site\_setup\_equlb.inp, site\_stup\_gcmc.inp, solv\_pert.inp, and solv\_stup\_eqlb.inp to be able to checkpoint during the MC and MD simulations.

We used the restart function native to CHARMM's MD module to restart the simulations from the last saved coordinates. However, maintaining the correct perturbations step over a single MD run required the script running CHARMM on the member devices to determine the

last step completed and generate a new .prt file with the modified steps to restart the perturbation at the correct step. Therefore, site\_pert.inp and solv\_pert.inp were modified to handle starting the MD simulation with a new .prt file.

We modified the site\_stup\_gcmc.inp to checkpoint at the end of each MC/MD cycle. The modified CHARMM input script generated .crd and .psf files for the system after each cycle. It then deleted and rebuilt the system anew for the next cycle. In the first-generation scripts, the cycles were continuous, allowing CHARMM to use the same water reservoir for all of the GCMC runs. In the second-generation scripts, the reservoir was rebuilt for each cycle.

#### **6.2.2.2 Performing MF-FEB Calculation**

We organized the MF-FEB calculations on World Community Grid into rescoring projects containing ~2,000 ligands that shared the same docking target. The projects corresponded to the virtual screening results of one of the DUD test sets or the virtual screening results in a *results.txt* file from Phase 1. Each job was given a four character ID. We divided virtual screening results into groups of 500 ligands, assigning each group an alphabet character (e.g., “a”). The ligands groups were further divided into work groups of 10 ligands labeled “g01” to “g50”. We submitted these work groups to World Community Grid similarly to how multiple docking simulations were bundled and submitted in Phase 1. Each ligand was renamed with a four character ID with the first character matching the ligand subset and the last three characters being a three digit integer (e.g., a001). We assigned the new names to the ligands file to avoid file name case sensitivity errors common in CHARMM. It then took six stages to build, process, and determine the results of a MF-FEB rescoring project. We automated the six stages using the *servoskull\_v02.py* and *techpriest\_v2.py* scripts and their dependent modules that are described in Section 6.2.2.3. This section describes the six stages.

#### 6.2.2.2.1 *Setup Stage*

The first stage, or setup stage, was performed on the project servers. We used the `servoskull_v02.py` script to build the work units for a rescoring project. The script required the directories and inputs described in Section 6.2.1 and a divided `results.txt` files corresponding to a ligand group (e.g., group “a”). The `servoskull_v02.py` script started setting-up the MF-FEB calculation by building directories corresponding to the organization method described above with each ligand in its own subdirectory. The script then copied and modified the `template.xml` file into the `target_ligand_charmm_param.xml` file containing the CHARMM parameters for the ligand-target system. Next, the script prepared the ligand by transforming the coordinate’s of the atoms in the `ligand.mol2` to match the docking pose, parameterizing the ligand, and centering the ligand on the origin as described in Section 2.4.1.1. The antechamber output and intermediate files generated during the preparation of the ligand were stored in the “`lig_wip`” directory and the final `ligand.prm`, `ligand.rtf`, and `ligand.crd` files were written to the “`compound`” directory. Next, the script copied the target files to the “`target`” directory, transposing the target to match the centered ligand as described in Section 2.4.1.2. The script would then build the prerun job as described below for submission to World Community Grid servers. The scripts concluded by generating the “`outhold`” directory to store the results of the following stages, and generating the `ligand_lig_rec.xml` file that tracked the progress of the MF-FEB calculations for the ligand.

#### 6.2.2.2.2 *Production Stages*

The next four stages, or production stages, involved performing the MF-FEB calculations on World Community Grid and were performed on project servers, World Community Grid servers, and member devices. The calculations were divided into four stages: prerun, Type A, Type B, and Type C. The stage name identified which stage of the MF-FEB calculations were run in a given job on a member device with each job being one CHARMM simulation. The prerun jobs corresponded with the setting up of the GSBP. The Type A jobs corresponded with

the GCMC simulations. The Type B jobs corresponded with equilibrating the systems. The Type C jobs corresponded with perturbation MD simulations. We also identified the Type B and Type C jobs with either the site or solvent (i.e., bulk) system. The Type C jobs were further divided by perturbation stage, the lambda step, and lambda direction. All jobs were identified with unique four character codes.

The process by which the `servoskull_v02.py` script would prepare a job was the same regardless of the stage. It would copy all files required to run the job into a directory to be archived into a work group `.tar.gz` file. The `target_ligand_charmm_param.xml` file contained the list of required files, which were obtained either from the “inputs” directory or in one of the sub-directories in the ligand’s directory. The `servoskull_v02.py` script would also copy and modify the corresponding template job xml file by adding the CHARMM command line options used on the member devices.

The overall process for performing the MF-FEB calculations worked as follows. All work was submitted to the IBM servers and received by the project servers in work units containing the jobs for 10 ligands. First, the setup stage generated the prerun work units. The IBM servers then received the preruns and ran them locally retaining the `target_mij.dat` and `target_phi.dat` binary files and returning the other files required to make the Type A jobs. With the preruns completed and returned, the Type As were built on the project server and submitted to World Community Grid servers. World Community Grid servers then added back the `target_mij.dat` and `target_phi.dat` files, distributed the Type A jobs to the member devices for the GCMC simulations, and returned the results to the project server which generated the Type B work units. Type B work units were submitted to World Community Grid, returned, processed into Type C work units, submitted, and returned again. The only files that were returned from the Type C jobs were the `result.wham` files needed to calculate the FEB. For each stage of the process, the `ligand_lig_rec.xml` would be updated to track the progress of the calculations.

#### 6.2.2.2.3 *Results Stage*

In the sixth stage, or results stage, the `servoskull_v02.py` script calculated the final FEB according the methods described in Section 2.4.2.3 and wrote the results to the `ligand_lig_rec.xml` file.

#### 6.2.2.3 *MF-FEB Job Creation and Management Scripts*

The section describes the Python scripts used on the project server, Ranger, to create and manage MF-FEB calculations on World Community Grid. Section 6.2.2.3.1 describes the master python scripts. Section 6.2.2.3.2 describes the “`dgtool_pro`” module that provided many of the functions and objects used by the master scripts.

##### 6.2.2.3.1 *Master Scripts*

The master scripts, used on the project server to manage the MF-FEB calculations for World Community Grid, were `techpriest_v02.py`, `servoskull_v02.py`, `arbitor_v01.py`, `archivist_01.py`, and `master_archivist.py`. This section describes these scripts.

###### 6.2.2.3.1.1 `techpriest_v02`

The `techpriest_v02.py` script built the launcher scripts used to submit `servoskull_v02.py` jobs on Ranger. This script served as the highest-level control script for managing the MF-FEB calculations. `Techpriest_v02.py` read the files from directories containing either the divided `results.txt` files from Phase 1 or returned work unit `.tar.gz` files from the different stages for the MF-FEB calculations. In addition to the input directories, the `techpriest_v02.py` script accepted command line options for job submission parameters for Ranger. The script then generated `job_setup_date.prm` and `job_setup_date.sge`, or `job_prodcution_date.prm` and `job_prodcution_date.sge` files depending upon if it was preparing jobs for the set-up stage or the production and results stages, respectively. The `.prm` files contained the command line calls for the `servoskull_v02.py` jobs to be run with one divided `results.txt` file or one returned work unit. The `.sge` files contained the Ranger job submission script.

#### 6.2.2.3.1.2 servoskull\_v02

The `servoskull_v02.py` script set-up and built the work units for the MF-FEB calculations as described in Section 6.2.2.2. In addition to command line options setting parameters for the MF-FEB calculation, the key `servoskull_v02.py` inputs were the “Set-up” or “Production” command line options and the corresponding `results.txt` files from Phase 1 or returned work unit `.tar.gz` files. Depending on the option selected and corresponding input files, the script would either set-up each ligand in the `results.txt` file generating setup work units, or process a work unit to the next stage (e.g., Type A to Type B, or Type C to results). The `servoskull_v02.py` script required the `dgtool_pro` module, described in Section 6.2.2.3.2, to process the inputs, and generate work units and results. The script generated the work unit’s `.tar.gz` files and the md5 checksum files for submission to World Community Grid servers and updated the `ligand_lig_rec.xml` file, recoding the work it had performed.

#### 6.2.2.3.1.3 arbitor\_v01.py

The `arbitor_v01.py` script generated reports on the progress of a rescoring project and gathered final FEB calculations. The script took three command line options, one indicating the name of the rescoring project, and one or both of the “progress” flags to generate a progress report, and/or the “results” flag to generate a results report. Using the directory organization generated by the `servoskull_v02.py` script, the `arbitor_v01.py` script would read the `ligand_lig_rec.xml` files to generate the reports. The progress report was printed to the screen and listed the number of ligand-target systems that had completed each stage of the MF-FEB calculations. The results report was written to `projectname_results_report.txt` file. The file was a tab delineated table with each line having the results of one ligand. The result lines started with ZINC name and new name for the ligands then contained either the FEB energy breakdown for successfully completed calculation, or “Failed” or “Incomplete” (sic) depending on the status of the MF-FEB calculations.

#### 6.2.2.3.1.4 archivist\_01.py and master\_archivist\_v01.py

The `archivist_01.py`, and `master_archivist.py` files archived a completed rescoring project. The `archivist_01.py` would GNU-zip all the files in a work unit group, and then tape archived and generated a md5 checksum for the individual ligand-target system directories. The `master_archivist.py` file generated the `.prm` and `.sge` files to run the `archivist_01.py` script for each work unit group for a single rescoring project on Ranger.

#### 6.2.2.3.2 “dgtools\_pro” Module

The `dgtools_pro` module was a collection of modules used by the master scripts to create and process ligand-target systems of MF-FEB calculations on World Community Grid. This section briefly describes each of the sub-modules and their key functions. The sub modules are: `arch.py`, `crdtools.py`, `dicxmltools.py`, `mol2dgprep.py`, `mol_obj.py`, `WU_build.py`, `WU_check.py`, `WU_prep.py`, and `WU_results.py`.

##### 6.2.2.3.2.1 arch.py

The `arch.py` module compressed the *result.wham* files to conserve hard drive space. The “arch” function compressed the files in a directory using GNU-zip. The “extract” function extracted the files in a directory.

##### 6.2.2.3.2.2 crdtools.py

The `crdtools.py` module contained the `rep_crd` object used to process the target CHARMM coordinate files to match a transposed ligand. An instance of the `rep_crd` object was called using a *target.crd* file. The objects key methods generated new *target.crd* files, “write\_crd”, and transposed the atomic coordinates, “center\_on”.

##### 6.2.2.3.2.3 dicxmltools.py

The `dicxmltools.py` module converted python dictionary objects to xml files, and xml files to Python dictionary objects. We used xml files to store parameters for and track the progress of the MF-FEB calculations. The “dic\_to\_xml” function converted python dictionary objects to xml files. The “xml\_to\_dic” function converted xml files to python dictionary objects.

#### 6.2.2.3.2.4 mol2dgpyp.py

The mol2dgpyp.py module converted *ligand.mol2* files into *ligand.rtf*, *ligand.prm*, and *ligand.crd* files for CHARMM using the “moltools\_mol2” object. The moltools\_mol2 object was called using a *ligand.mol2* and contained methods for manipulating the ligand. The “save\_mol2” method wrote a new *ligand.mol2* file. The “rename\_atoms\_unique” method renamed the atoms in the ligand assigning them unique names. The “charmm\_prep” method generated the CHARMM ligand files using antechamber. The “geocenter” method transformed the ligand’s atomic coordinates to the origin returning the information to transpose the target.

#### 6.2.2.3.2.5 mol\_obj.py

The mol\_obj.py script contained the “mol” object that generated a *ligand.mol2* file from the results pose of a docking simulation from Phase 1. The mol object was called using the *ligand.mol2* file from the ZINC database. The “build\_tors” method used the *ligand.pdbqt* file used for Phase 1 docking simulations to determine the torsion bonds in the ligand. The “tran” and “va\_rot” methods used the pose information to transform the coordinates in the *ligand.mol2* file into the docking pose. The “write\_mol2” method wrote the transformed coordinates to a *ligand.mol2* file.

#### 6.2.2.3.2.6 WU\_build.py

The WU\_build.py module prepared the Type A, Type B, and Type C work units. The “GCMC” function built the Type A work units. The “system\_eq1” function built the Type B work units for the site and solvent systems. The “pert\_repu” function built the Type C VdW repulsion work units for the site and solvent systems. The “pert\_qdc” function built the Type C VdW dispersion and electrostatic interaction work units for the site and solvent system, and the Type C positional constraints work units for the site system. The “site\_pert\_rmsd” function built the Type C conformational constraints work units for the site system.

#### 6.2.2.3.2.7 WU\_check.py

The WU\_check.py module contained the “crd\_coor\_check” function that checked a .crd file’s coordinates. The function returned true if the coordinates were greater than 100 Å from the origin or where “NaN” coordinate indicted an error. The module was designed to be the location of future error checking functions.

#### 6.2.2.3.2.8 WU\_prep.py

The WU\_prep.py module performed the setup stage for MF-FEB calculations on World Community Grid. The “ligand” function prepared the ligand files. The “rep” function prepared the target files. The “charmm\_xml\_param” function generated the *target\_ligand\_charmm\_param.xml* file. The “prep\_GSBP” function built the prerun work units.

#### 6.2.2.3.2.9 WU\_results.py

The WU\_results.py module calculated the FEB from the Type C work unit results. The “get\_feb” function called the following functions to determine the energy components for the FEB, calculate the FEB, and return a python dictionary object containing the results. The “num\_int\_par\_dx” function computed the area under a curve using Simpsons’ Rule for numerical integration. The “qdc\_FEB” function calculated the energy from the perturbation of the electrostatic interactions, VdW dispersion for the site and solvent systems, and the releasing of the positional constants using WHAM analysis. The “rep\_FEB” function calculated the energy from the perturbation VdW repulsion for the site and solvent systems using WHAM analysis. The “rmsd\_FEB” function calculated the energy from releasing of the conformation constants using WHAM analysis and equation 1-22. The “const\_FEB” function calculated the energy from the restrains by solving equation 1-23. The “lrc\_FEB” function read in the long-range VdW correction factor from the prerun outputs.

### **6.2.3 MF-FEB Grid Performance**

This section examines the performance of MF-FEB calculations on World Community Grid. Section 6.2.3.1 looks at runtimes, and Section 6.2.3.2 discusses the failure rate of MF-FEB calculations on World Community Grid

#### ***6.2.3.1 Runtime***

On Ranger, a MF-FEB calculation for one ligand-target system took a wall-clock time of about 72 hours, with minimal queue times and 64 processors. For between four and sixteen simultaneously MF-FEB calculations, it took a week to complete all the calculations. On World Community Grid, a MF-FEB calculation for one ligand-target system took maximum wall-clock time of about 66 hours to run on about 170 of the member computers, not including the time required sending and receiving jobs from World Community Grid servers and generating jobs on the project server. The full wall-clock time to rescore the TRYP test set took a total of 142 days. With some overlap, it took 37 day to run preruns, 27 days to run Type A work units, 21 days to run Type B work units, and 64 days to run Type C work units and final FEB determination. However, throughout the 142 days World Community Grid performed more than 1,500 MF-FEB calculations. Therefore, as discussed in Section 6.1.3, the advantage of using a grid is throughput and not efficiency.

#### ***6.2.3.2 Job Failure Rates***

Some MF-FEB calculations were prone to fail and required monitoring and troubleshooting to insure completion. One of the challenges of implementing the MF-FEB calculation on World Community Grid was that we could not closely monitor the MF-FEB calculations for two reasons. First, member devices performing the calculations returned only the minimal information, preventing us from examining the CHARMM output file and the MD trajectories files commonly used for troubleshooting. Second, there were a large number of calculations, making fixing and resubmitting failed jobs very time consuming. Because a majority of the MF-FEB calculations ran successfully, we elected not to attempt to troubleshoot

failed jobs and treated failed jobs as non-binders. However, the rescoring of the ERAG and TRYP test sets described in Chapter 7 provided an opportunity to benchmark the failure rates of the MF-FEB calculation on World Community Grid.

Overall, 82% of the ERAG test sets and 71% of the TRYP test sets MF-FEB calculations successfully completed. Figure 6-1 and Figure 6-2 show the cumulative success and failure rates for the ERAG and TRYP test sets respectively. Very few ligand-target systems failed during the setup or prerun stages. Failures in the setup stage were likely due to errors arising from the parameterization of the ligand with antechamber. Prerun failures were likely due to errors in the generation of the GSBP. A few percent of the jobs were lost during the Type A and Type B stages, with most being lost in Type A corresponding with errors arising from the first time simulations were performed on the system. The largest loss occurred during the Type C stage corresponding to the perturbation runs, with more site runs failing than solvation runs. These failure rates correspond with the failures observed during MF-FEB calculations using the TACC supercomputers. Most errors seen at TACC were due to errors in MD simulations caused by erroneous atom movement or placement that caused unrealistic interactions and energies that would crash the simulation. The total failure rates were a combination of calculations where either the site or the solvation simulations failed, or WHAM analyses that did not converge or failed for technical reasons.

We looked for trends in the success and failure of the MF-FEB calculations to identify any possible causes of failure, and to find a metric that might indicate a failure. Three percent of the failed MF-FEB calculations in the TRYP test set were binders corresponding to the percentage of binders in the total population, showing no indication of binder or non-binder bias on job success. However, none of the ERAG test set binders failed. The ERAG binders were the first set of compounds tested in the MF-FEB simulations and were sent to reliable member devices suggesting that some of the failure rate may be due to computer performance. Additionally, we analyzed the ligand databases using ChemMine (<http://chemmine.ucr.edu/>) to

generate molecular descriptors for the ligands to determine if there was any correlation between any of the descriptors and the success or failure of the MF-FEB calculations. Correlation between a molecular descriptor and the success or failure might indicate a potential limitation of MF-FEB calculations; however, we saw no correlation with any of the descriptors.

An average failure of 25% is high, but not surprising considering the nature and complexity of the calculations. Improving the MM code should increase stability of the MD simulation, and improve the overall performance of the MF-FEB calculation on member devices; thereby, reducing the failure rate.

Figure 6-1 – MF-FEB calculation cumulative success and failure rates by stage on World Community Grid for ERAG test set.

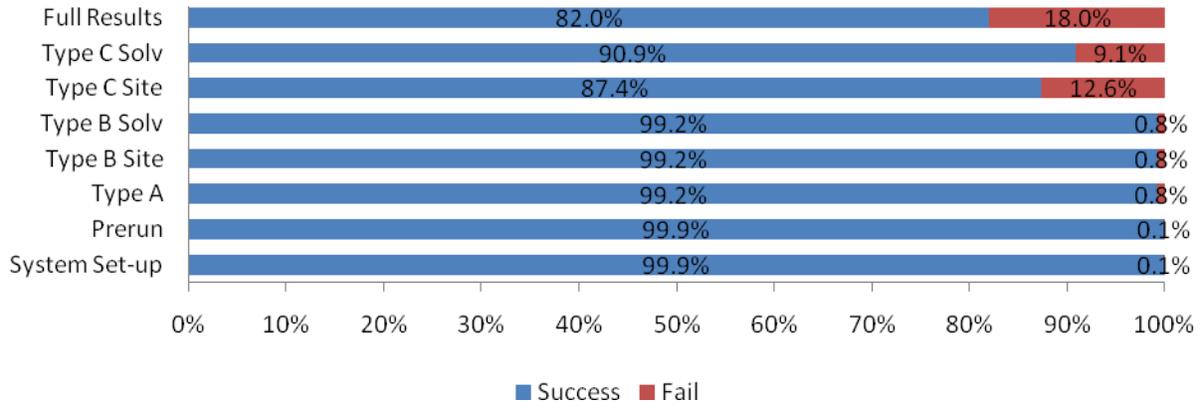
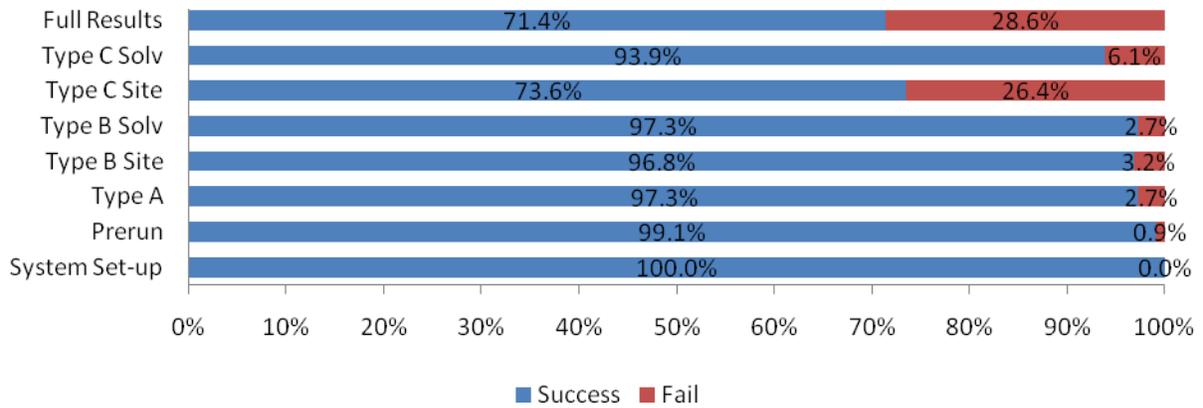


Figure 6-2 – MF-FEB calculation cumulative success and failure rates by stage on World Community Grid for TRYP test set.



## **Chapter 7 Efficiency and Efficacy of Rescoring Docking Generated Poses by Perturbation Free Energy of Binding Calculation in Large Datasets**

The goal of this work was to determine if rescoring docking generated poses using the MF-FEB calculations would improve enrichment in virtual screening results over docking scoring functions. In Chapter 4, we showed that MF-FEB calculations could use docking generated poses for FEB predictions. Additionally, we saw improved enrichment in virtual screening of a 30 compound chemical database and significant discrimination between binder and non-binders when we rescored AutoDock4 generated pose with MF-FEB calculations. While the results in Chapter 4 supported our hypothesis, we only tested the MF-FEB rescoring method in the L99A T4 lysozyme test system, due to limited computational resources. We therefore needed to test MF-FEB rescoring using larger (>1,000 compounds) and more drug-like test sets in order to more fully test our hypothesis and see if MF-FEB rescoring could be used effectively as a drug discovery tool. We required formidable computational resources to perform MF-FEB rescoring for thousands of compounds, which we obtained through World Community Grid, as described in Chapter 5. Additionally, DUD provided a collection of virtual screening test sets that Cross et. al. had used to test other docking programs, thereby allowing us to compare MF-FEB rescoring to other docking programs and scoring functions. (Huang, Shoichet et al. 2006; Cross, Thompson et al. 2009) With the DUD test sets and the computational power of World Community Grid, we were able to test MF-FEB rescoring on larger drug-like test sets.

In this chapter we examine if MF-FEB rescoring can improved virtual screening enrichment over AutoDock4's scoring function in large drug-like test sets. We selected two test sets from the DUD database, estrogen receptor agonist (ERAG) and trypsin (TRYP), which we had used to determining virtual screening result filters as described in Section 3.1.2. We virtually screened both test sets with AutoDock4 and rescored the docking generated poses using MF-FEB calculations. Because each test set contained ~2,000 compounds we were able to rescore every compound, binder and decoy, in each test set allowing for a comparison with the

results of Cross et al. (Cross, Thompson et al. 2009) In both the ERAG and TRYP test sets MF-FEB rescoring improved enrichment over AutoDock4's scoring function. However, MF-FEB calculations did not outperform all docking scoring functions and therefore needed to be improved in order to be employed as a drug discovery tool.

This chapter is divided into three sections. Section 7.1 describes the methods employed to virtually screen and rescore the ERAG and TRYP test sets. Section 7.2 discusses the virtual screening enrichment due to MF-FEB rescoring in both test sets. Section 7.3 concludes the chapter by discussing insights into the performance of the MF-FEB calculations obtained from the two virtual screening experiments.

## **7.1 METHODS**

This methods section is divided into three sections. Section 7.1.1 describes the ERAG and TRYP DUD test systems. Section 7.1.2 describes the virtual screening methods. Section 7.1.3 describes the MF-FEB rescoring methods.

### **7.1.1 DUD Test Sets**

The ERAG and TRYP test sets were the same we used in Section 3.1.2.1 and are described in that section.

When considered as test sets for MF-FEB rescoring trials, ERAG and TRYP tests different aspects of the MF-FEB calculations. ERAG provided a test set that was similar to the L99A T4 lysozyme test set used in Chapter 4 because the ligands in the test set were hydrophobic and the binding site has little contact with solvent waters. However, ERAG binders were larger and more drug-like than those of the L99A T4 lysozyme test and were dependent on a key H-bonding interaction for binding. TRYP provided a different test set from the L99A T4 lysozyme test set because the ligands of the TYRP test set were hydrophilic and charged, with binding being dependent on electrostatic interaction, and a binding site that was largely solvent

exposed. Therefore, these two systems provided contrasting test sets with which to prove the MF-FEB calculations.

The main disadvantage of the DUD systems was that the decoy compounds were selected from the ZINC database based on their chemical similarity to the known binders in the test set and not from experimental results. Therefore, it was possible that a few of the decoys may be true binders that had never been tested.

### **7.1.2 Virtual Screening**

We performed the virtual screening of the ERAG and TRYP test sets using the methods described in Sections 2.3 and 3.1.2.1 on TACC's Lonestar supercomputer.

### **7.1.3 MF-FEB Rescoring**

We rescored the docking generated poses on World Community Grid as part of Phase 2 of DDDT using the methods described in Sections 2.4.2 and 6.2. In rank ordering the rescoring results, we retained failed MF-FEB runs and placed them at the end of the rankings to mimic how failed results are treated as non-binders in a prospective virtual screening project.

## **7.2 ENRICHMENT**

In determining the effect of MF-FEB rescoring on the virtual screening results, we utilized three analysis methods. First, we looked at the ability of AutoDock4's scoring function and MF-FEB calculations to discriminate between binders and decoys using two-tailed Student T-tests for populations of unequal variance. Second, we plotted the enrichment curves of each scoring method. Figure 7-1 and Figure 7-2 contain the enrichment curves for AutoDock4's scoring function and MF-FEB rescoring for ERAG and TRYP test sets respectively. Finally, we examined potential experimental impact of virtual screening results by determining success rate.

When, virtual screening results are validated experimentally, they are often tested by rank order with the top ranked compound being tested first. Therefore, the success rate in bench top

testing would be the percentage of true binders in the tested compounds. In a retrospective study, success rate corresponds to the percentage of known binders in the first “n” compounds in the list of rank ordered virtual screening results. For the ERAG and TRYP virtual screening studies, we calculated and plotted the success rate for the first 10, 20, n+10 ... 100 compounds. Figure 7-3 and Figure 7-4 contain the success plots for AutoDock4’s scoring function and MF-FEB rescoring for ERAG and TRYP respectively. Success rate plots are similar to early enrichment plots and ROC plots, but casts enrichment in terms of hypothetical experimental results.

For the ERAG test set, rescoring the ME-FEB calculations provide a significant increase in enrichment over AutoDock4’s scoring function. Both AutoDock4’s scoring function and MF-FEB calculations were able to discriminate between binders and decoys. The mean docking scores were  $-6.74 \pm 0.03$  kcal mol<sup>-1</sup> for the decoys and  $-7.4 \pm 0.2$  kcal mol<sup>-1</sup> for the binders. The mean FEB from the MF-FEB calculations were  $-9.4 \pm 0.2$  kcal mol<sup>-1</sup> for the decoys and  $-17.1 \pm 1.5$  kcal mol<sup>-1</sup> for the binders. While the docking score was able to significantly discriminate between binders and decoys with a p-score of  $3 \times 10^{-7}$  and a difference in means of  $\sim 1$  kcal mol<sup>-1</sup>, the MF-FEB calculations improved discrimination with a p-score of  $4.6 \times 10^{-14}$  and a difference in means of  $\sim 8$  kcal mol<sup>-1</sup>. As shown in Figure 7-1, both AutoDock4’s docking scoring function and MF-FEB virtual screening calculations enriched the test set for binders better than random selection, with MF-FEB rescoring improving enrichment over AutoDock4’s scoring function. The success rate, shown in Figure 7-3, for the MF-FEB calculations was more than two times greater than the docking scoring function. In the first ten compounds, MF-FEB rescoring identified five binders while the docking scoring function only identified two. Overall, the enrichment results for the ERAG test set were similar to those of the small L99A T4 lysozyme test set in that MF-FEB rescoring significantly improved discrimination and enrichment over AutoDock4’s scoring function. However, the MF-FEB method did generate false-positives.

For the TRYP test set, MF-FEB rescoring improved enrichment over AutoDock4's scoring function, but the overall virtual screening and rescoring were not able to enrich the test set binders. The mean docking scores were  $-7.74 \pm 0.08$  kcal mol<sup>-1</sup> for the decoys and  $-5.1 \pm 0.8$  kcal mol<sup>-1</sup> for the binders. The mean  $\Delta\Delta G_{\text{binding}}$  from the MF-FEB calculations were  $-3.0 \pm 0.7$  kcal mol<sup>-1</sup> for the decoys and  $1.1 \pm 4.8$  kcal mol<sup>-1</sup> for the binders. For both scoring methods, the mean FEB of the binders was incorrectly higher than the mean for the decoys. While AutoDock4's scoring function could discriminate between binders and decoy (p-score of  $6 \times 10^{-8}$ ), albeit inversely, the standard deviation of  $\Delta\Delta G_{\text{binding}}$  predicated by MF-FEB calculations was too large to discriminate binders and decoys (p-score of 0.1). Consistent with the discrimination results, AutoDock4's scoring function did not enrich the compound database for binders, but did the opposite moving binders to the bottom of the ranked ordered compound database as seen in Figure 7-2. However, MF-FEB rescoring improved enrichment, enriching the database slightly better than random for the first ~10% of the test set (Figure 7-2). Due to the negative enrichment by AutoDock4's scoring function, the success rate for the first 100 compounds was zero as shown in Figure 7-4. MF-FEB rescoring did identified five binders in the first 100 compounds (Figure 7-4). Overall, MF-FEB rescoring did increase enrichment in the TRYP test set and generated "hits" which AutoDock4's scoring function did not. However, the MF-FEB calculations were predicting high and even positive  $\Delta\Delta G_{\text{binding}}$  for a number of binders indicating that the MF-FEB calculations need to be improved as discussed in Section 8.3.

Because we utilized the DUD test sets, we were able to compare the enrichment of MF-FEB rescoring to other docking scoring methods. Figure 7-5 and Figure 7-6 are enrichment curves generated for the data kindly provided by Dr. Cross corresponding to the enrichment curve for different docking programs that appeared in Cross et al. (Cross, Thompson et al. 2009) overlaid with our MF-FEB enrichment curve for ERAG and TRYP respectively. The goal of these figures was not to compare MF-FEB rescoring to any single scoring function but to compare the rescoring method to the field in general. As shown in Figure 7-5, for the ERAG test

set, MF-FEB rescoring was one of the top enrichment methods, but other docking scoring functions outperformed MF-FEB rescoring. As shown in Figure 7-6, for the TRYP test set many docking scoring functions outperformed MF-FEB rescoring. Overall, MF-FEB rescoring is comparable to other scoring functions in performance, and did not distinguish itself as a drug discovery tool, in that MF-FEB rescoring could enrich databases, but its performance was system dependent and generated false positives.

The goal of this section was to determine if MF-FEB rescoring can improve virtual screening enrichment over AutoDock4's scoring function in large and drug-like test sets. In both the ERAG and TRYP test set, MF-FEB rescoring did improve enrichment and significantly increased the odds of getting hits for the virtual screening results. However, MF-FEB rescoring did not outperform other docking programs and generated false positives; therefore, it needs to be improved before being used as a drug discovery tool. See Section 8.1 for a discussion of the utility of MF-FEB rescoring as a drug discovery tool.

Figure 7-1 – Plot of DUD’s estrogen receptor agonist test set enrichment curves sorted by AutoDock4’s scoring function and MF-FEB rescoring. Dashed lines correspond to ideal enrichment and random enrichment.

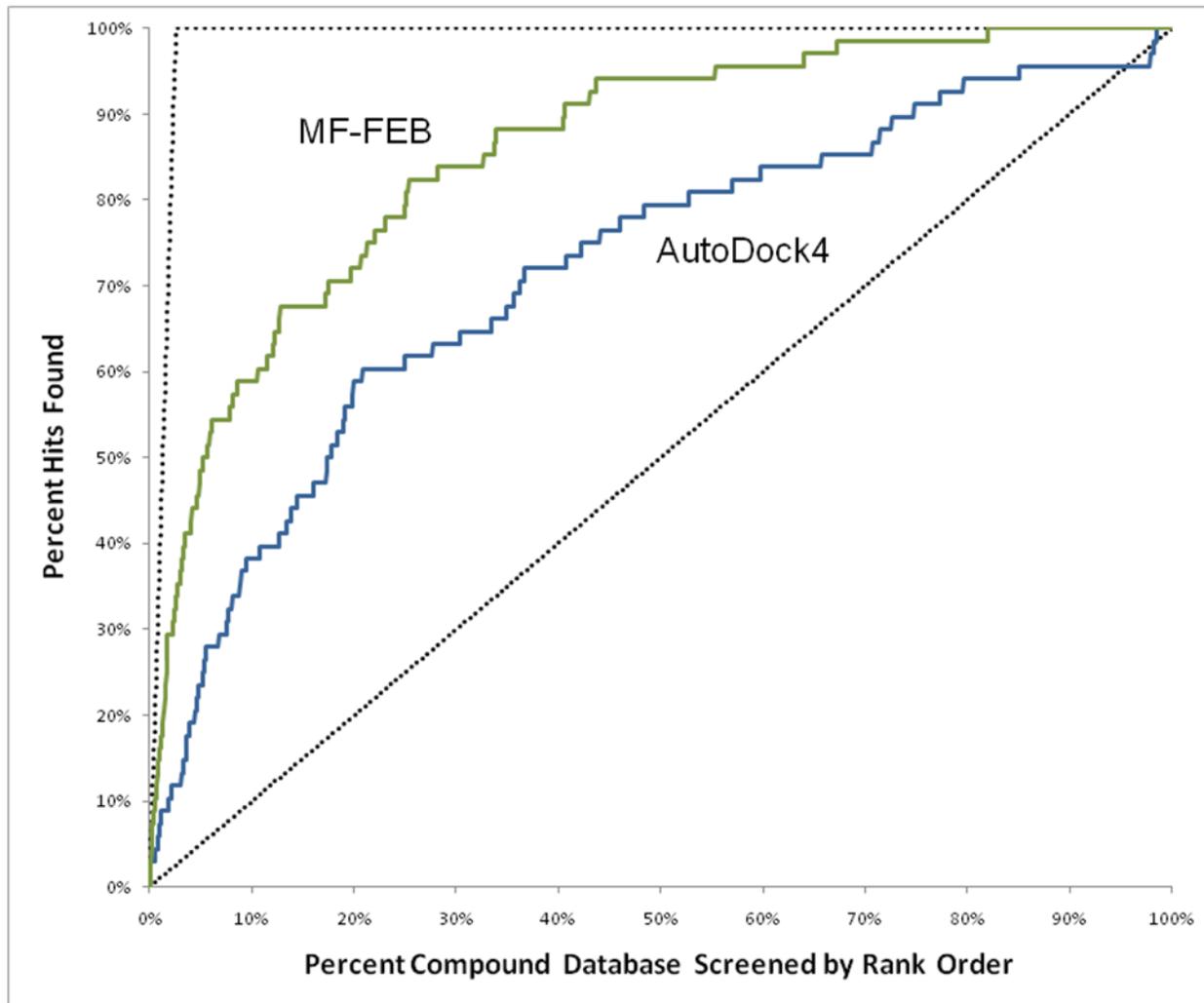


Figure 7-2 – Plot of DUD’s trypsin test set enrichment curves sorted by AutoDock4’s scoring function and MF-FEB rescoring. Dashed lines correspond to ideal enrichment and random enrichment.

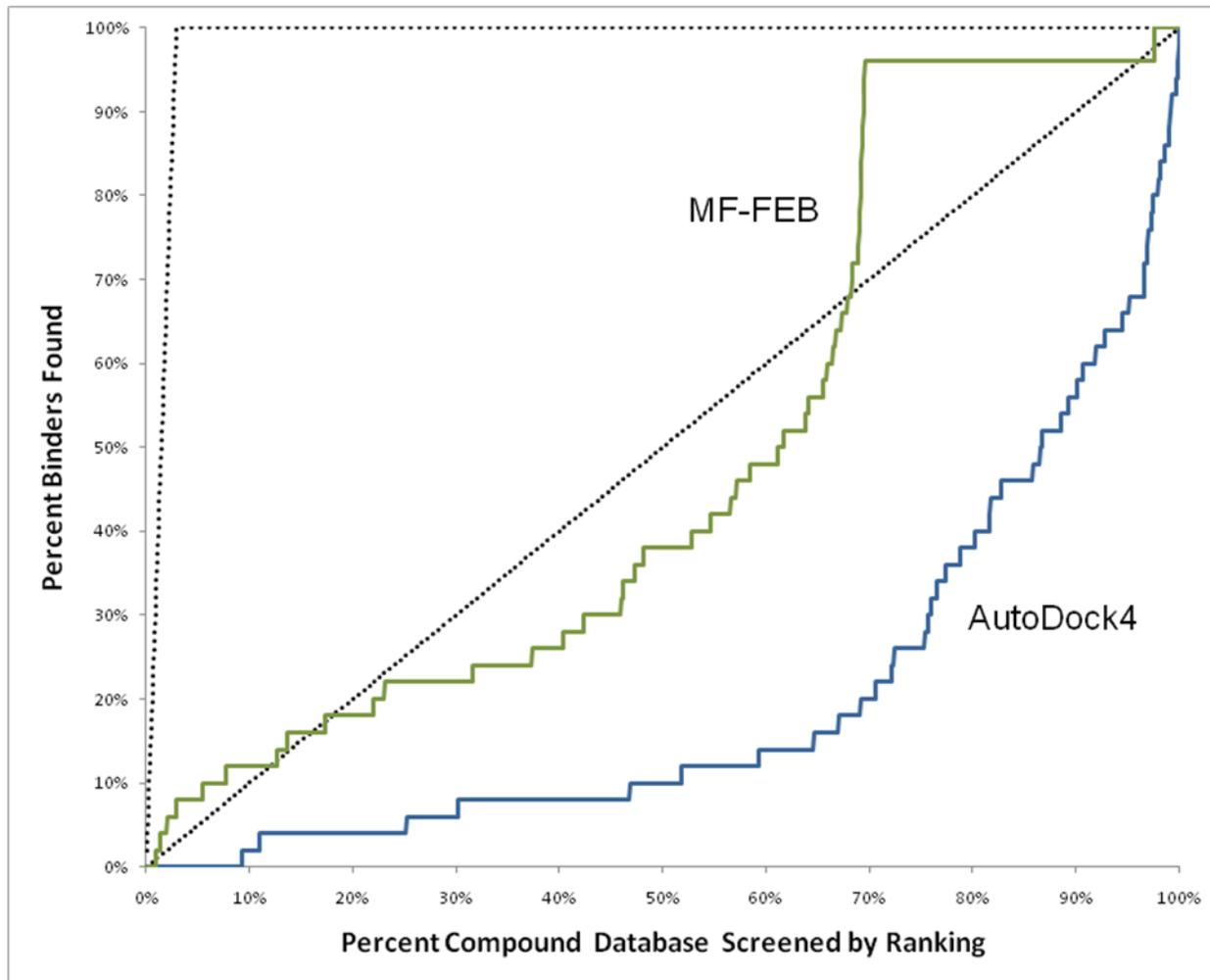


Figure 7-3 – Success rates plot for DUD’s estrogen receptor agonist test set. All bars start at 0% and are overlaid for comparison. Ideal corresponds to the total number of binders in the test set.

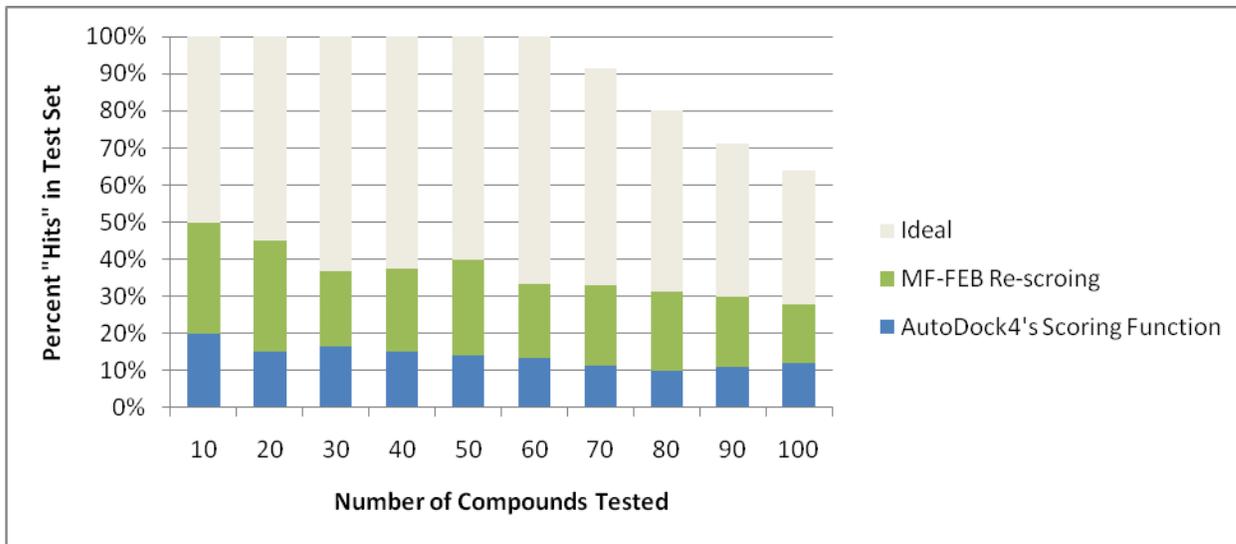


Figure 7-4 – Success rates plot for DUD’s trypsin test set. All bars start at 0% and are overlaid for comparison. Ideal corresponds to the total number of binders in the test set. There were no binders found using AutoDock4’s scoring function.

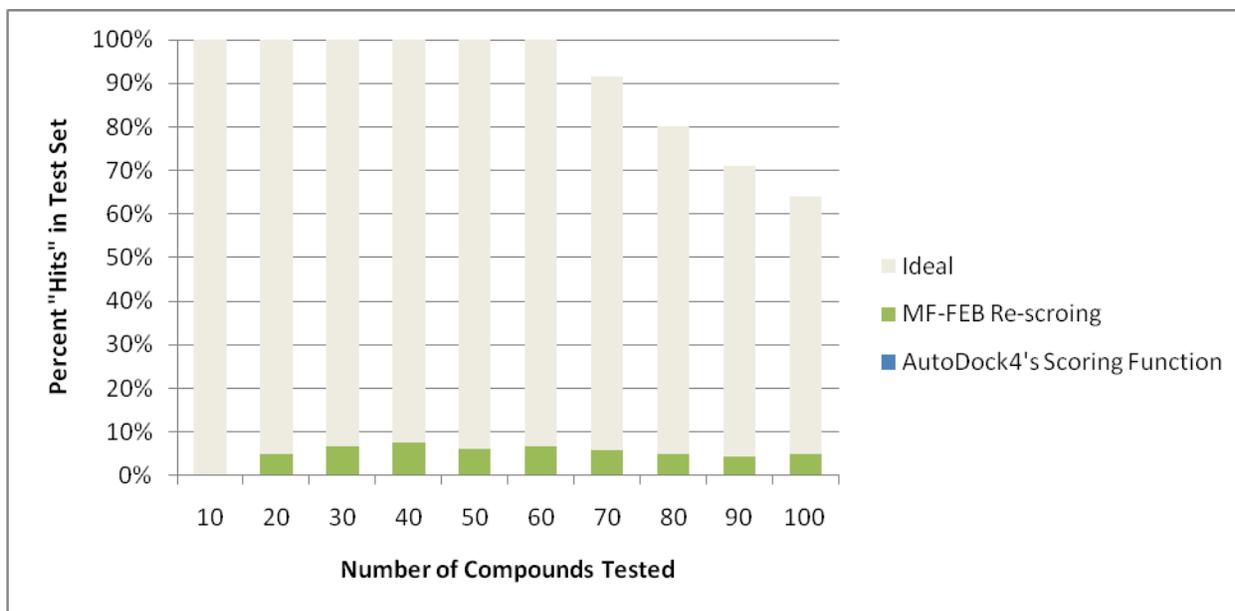


Figure 7-5 – Overlay of MF-FEB rescoring enrichment curve on enrichment curves from other docking programs for DUD's estrogen receptor agonist test set. (Cross, Thompson et al. 2009)

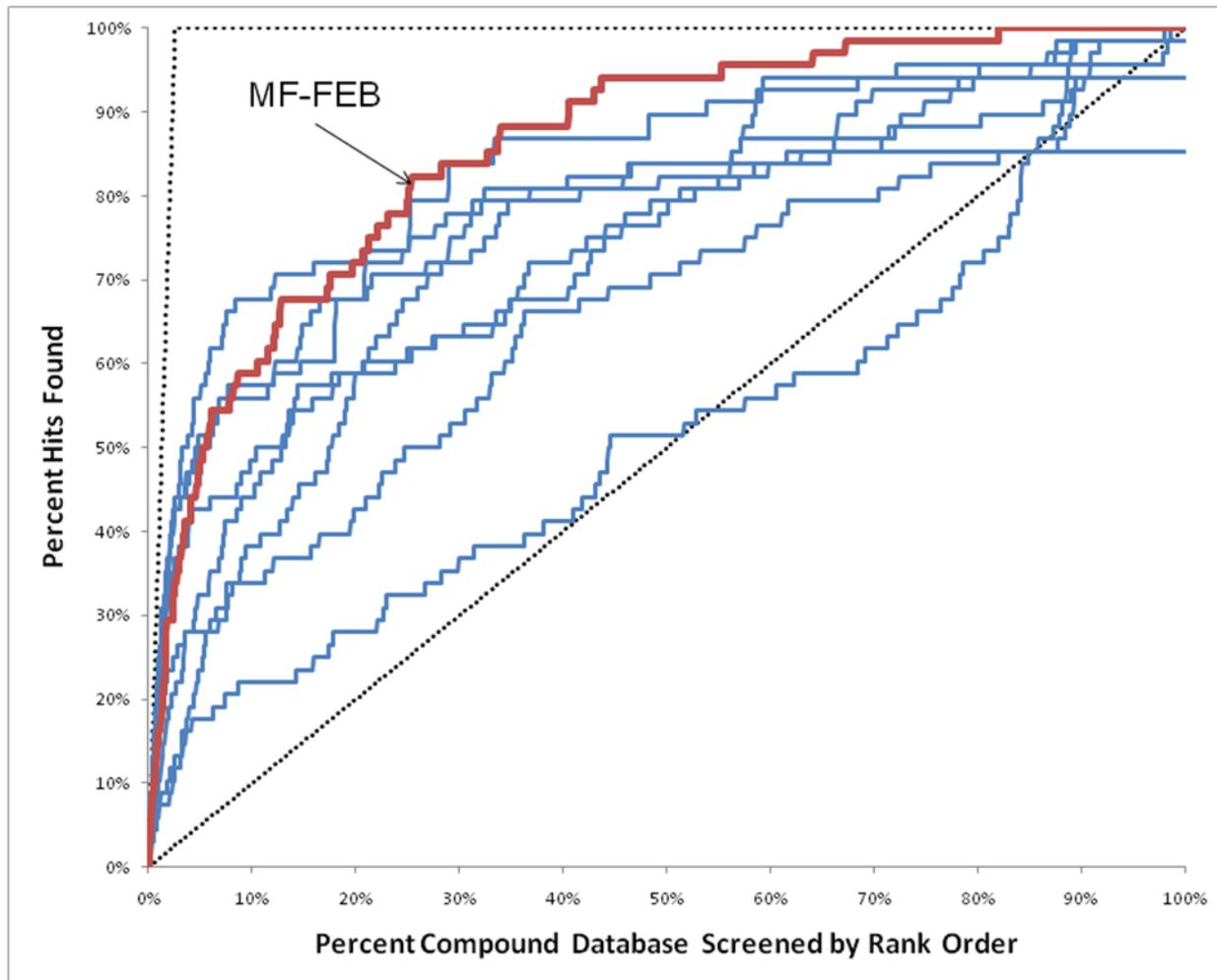
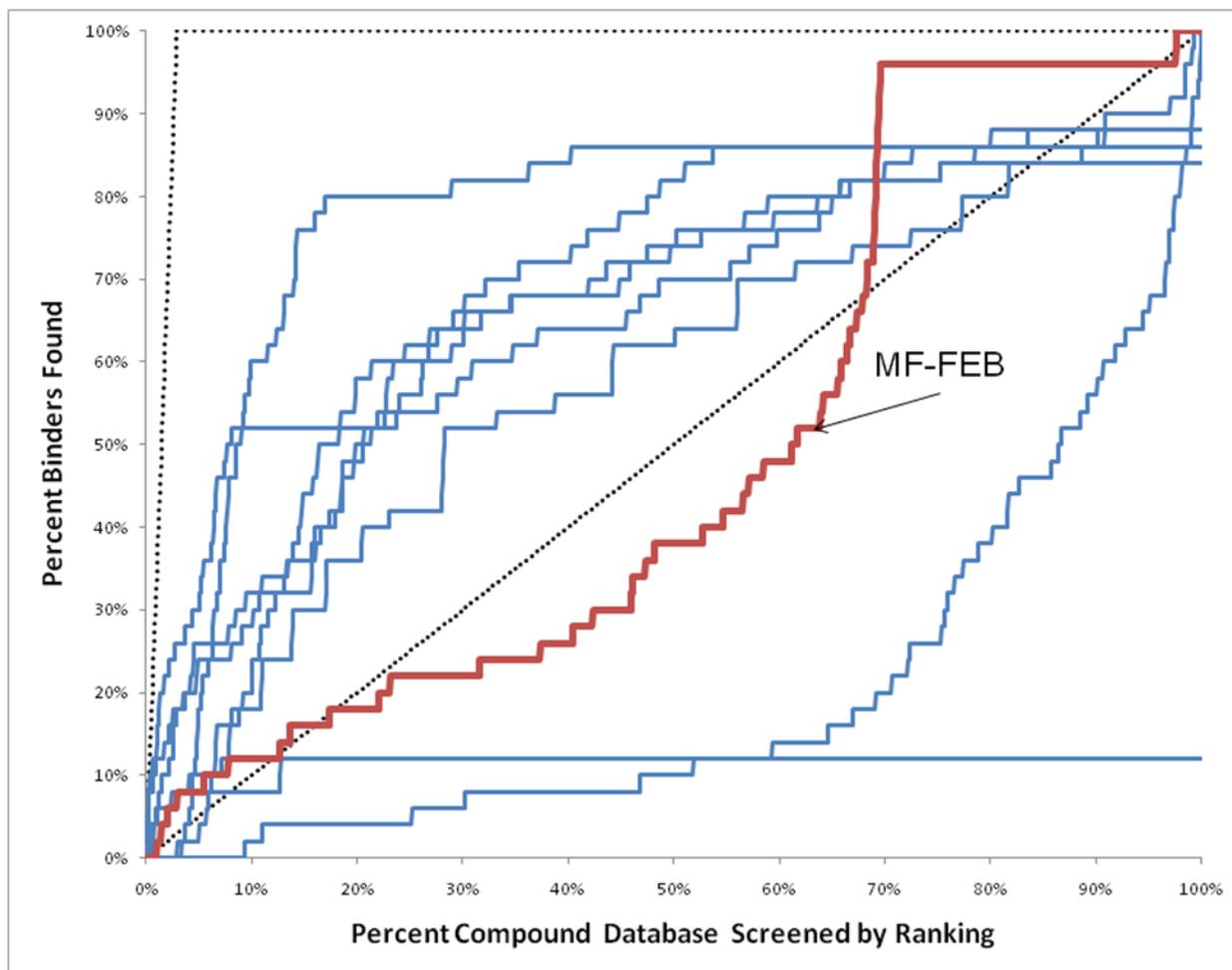


Figure 7-6 – Overlay of MF-FEB rescoring enrichment curve on enrichment curves from other docking programs for DUD's trypsin test set. (Cross, Thompson et al. 2009)



### 7.3 SOURCES OF ERROR IN THE MF-FEB CALCULATIONS

MF-FEB rescoring's average performance relative to other docking programs prompted us to make initial inquiries into how the MF-FEB calculations could be improved. The first question was what influence did pose have on  $\Delta\Delta G_{\text{binding}}$  and final ranking? The second question was could trends in the component energy terms that comprised the FEB prediction identify a way to improve the MF-FEB calculations? We addressed the second question in two ways. To begin, we examined the ability of the component energy terms to discriminate between binders and decoys. Then, we looked at the correlations between the component energy terms. This section addresses these two questions. Section 7.3.1 examines the effect of pose on ranking. Section 7.3.2 describes how component energy terms discriminate between binders and decoys, and the correlations between the component energy terms.

#### 7.3.1 Pose and Ranking

In comparing the improvement in MF-FEB rescoring enrichment in the TRYP test set to the enrichment in the ERAG test set, it appeared that the results of the docking might have influenced the rescoring. The MF-FEB rescoring improved enrichment in both cases, but in TYRP it only improved negative enrichment to random levels of enrichment (i.e., junk in, junk out). Additionally, as we compared MF-FEB enrichment from AutoDock4 results to other docking programs, we asked to what degree MF-FEB rescoring might improve the enrichment of results from other docking programs? Ideally, the way to test whether MF-FEB rescoring would improve other docking results would be to perform a new virtual screening with a different docking program or scoring function and rescore the results. However, the only information passed from the docking program to the MF-FEB calculations was the ligand pose. In Chapter 4, we showed that docking poses could be used for MF-FEB calculations, but the trends of  $\Delta\Delta G_{\text{binding}}$  predictions to RMSD suggest that accurate  $\Delta\Delta G_{\text{binding}}$  predictions were obtained from those poses that resemble the co-crystal pose. Therefore, instead of performing a new virtual screening project, we asked the question of how the  $\Delta\Delta G_{\text{binding}}$  prediction from MF-FEB

calculations would compare between the self-dock and co-crystal ligand poses of the target structures used in the ERAG and TRYP test sets. By testing the co-crystal structure, we were using the best possible docking result.

Each DUD test set has a co-crystal target protein structure whose ligand we included in the test sets for virtual screening. Table 3-6 contains the target PDB structures codes for the ERAG and TRYP test sets. We performed MF-FEB calculations on the co-crystal structures of ERAG and TRYP targets on Ranger using the methods described in Section 2.4. We obtained the AutoDock4 self-docked results from the virtual screenings described in Section 7.2, although we needed to re-run the TRYP self-dock MF-FEB calculations on Ranger because they failed on World Community Grid. The results of the MF-FEB calculations for both the self-docked and co-crystal poses for ERAG and TRYP are in Table 7-1.

Table 7-1 –  $\Delta\Delta G_{\text{binding}}$  and rank order from MF-FEB calculations using different starting poses.

| System                     | Docking Pose |   |            | Co-Crystal Pose   |            |
|----------------------------|--------------|---|------------|---|------------|
|                            | RMSD (Å)     | $\Delta\Delta G_{\text{binding}}$ (kcal mol <sup>-1</sup> ) | Rank       | $\Delta\Delta G_{\text{binding}}$ (kcal mol <sup>-1</sup> ) | Rank       |
| Estrogen Receptor Agonists | 0.69         | -23.8   | 20 of 2638 | -21.09  | 46 of 2638 |
| Trypsin                    | 3.46         | -22.31  | 31 of 1714 | -27.23  | 16 of 1714 |

With ERAG, the difference between the self-docked and co-crystal ligand poses was a RMSD of 0.67 Å. The pose difference changed the predicted  $\Delta\Delta G_{\text{binding}}$  by ~2 kcal mol<sup>-1</sup>, comparable to energy differences between repeated runs of the same system as seen in Section 4.3. The difference in energy placed the co-crystal pose in the top 2% of the test set, while the

lower scoring self-docked pose was in the top 1%. Therefore, in the ranking of the ERAG test set the difference between the two poses was minor.

With TRYP, the difference between the self-docked and co-crystal ligand poses was a RMSD of 3.46 Å. The change corresponded to a repositioning to the portion of the ligand that extends out the P1 site. The co-crystal pose decrease the calculated  $\Delta\Delta G_{\text{binding}}$  by  $\sim 5 \text{ kcal mol}^{-1}$ , moving the targets ligand from the first 2% to the first 1% of the rank ordered test set. Therefore, in the TRYP test set using the co-crystal pose lead to an improved score and a minor improvement in rank order.

If poor poses was the cause of the modest enrichment in the virtual screening of the TRYP test set, then the co-crystal poses should have been at the very top of the rank order. However, in both ERAG and TRYP, using the co-crystal pose only had a minor effect in ranking; with ERAG co-crystal pose actually lowered in rank. Therefore, while a good pose was important to an accurate MF-FEB calculation as shown in Section 4.2, the average performances of the MF-FEB calculations was most likely due to inaccurate  $\Delta\Delta G_{\text{binding}}$  predictions.

### 7.3.2 The Relationships between MF-FEB Energetic Components

In both the ERAG and TRYP test sets, MF-FEB rescoring generated false positives and in the case of TYRP, MF-FEB rescoring was unable to discriminate between binders and non-binders. As indicated from the result in Section 7.3.1, improving MF-FEB calculations meant improving  $\Delta\Delta G_{\text{binding}}$  predictions themselves and not the docking results. The MF-FEB calculations are very complex and improving them could require modifying the run parameters, changing the force field and partial charge assignment, modifying the underling theory, or improving methods used to predict the FEB. With data from two virtual screening projects, we ask if the data contained any indication of what needed to be improved. In the MF-FEB calculations, the  $\Delta\Delta G_{\text{binding}}$  was the sum of energy components (see Section 1.4.4). Therefore, we examined each of these energy components looking at the relationships between them and

their ability to discriminate between binders and decoys. Identifying which energy components were aiding discrimination and determining their interdependence would indicate how we might improve the MF-FEB calculations

We looked at the discrimination of the different FEB energy components between binders and decoys by determining the mean of each population and comparing the binder and decoy populations using a two-tailed Student t-test. Table 7-2 and Table 7-3 contain the means for binders and decoys, and the p-score from the Student's t-test for each  $\Delta\Delta G_{\text{binding}}$  energy component for the ERAG and TRYP test set respectively. We examined the relationship between the  $\Delta\Delta G_{\text{binding}}$  energy components by determining the correlation between each of the  $\Delta\Delta G_{\text{binding}}$  energy components, treating the binders and decoys as a single population. Table 7-4 and Table 7-5 are heat maps based on the correlation values for the ERAG and TRYP test sets respectively.

MF-FEB calculations were able to discriminate between binders and decoys in the ERAG test. However, the mean of the  $\Delta\Delta G_{\text{binding}}$  was lower than expected if the results were accurately predicting  $\Delta\Delta G_{\text{binding}}$ . For binders, the mean  $\Delta\Delta G_{\text{binding}}$  for ERAG binders was  $-17 \text{ kcal mol}^{-1}$ , which would correspond to the average ERAG binder being a picomolar binder. The mean  $\Delta\Delta G_{\text{binding}}$  for the decoys was  $-9 \text{ kcal mol}^{-1}$ , which would correspond to the average decoy being a micromolar binder. In comparing the contributions to the  $\Delta\Delta G_{\text{binding}}$  for the free energy of solvation and the  $\Delta G_{\text{site}}$ , as with the L99A T4 lysozyme system, the  $\Delta G_{\text{solv}}$  contributed more to discrimination than the  $\Delta G_{\text{site}}$ .  $\Delta G_{\text{solv}}$  was  $\sim 5 \text{ kcal mol}^{-1}$  less favorable for the binders than the decoy largely due to less favorable electrostatic interaction. For the  $\Delta G_{\text{site}}$ , favorable VdW interactions drove the discrimination of the binders and the decoys. Additionally, there was a small but highly significant discrimination due to the higher energy required to restrain the ligand's position for the decoy, suggesting, with increased VdW repulsion, that the decoys did not fit the binding site as well as the binders. These trends in energy components were consistent with ERAG binders being hydrophobic molecules. In examining the correlations between different energy components, we observed that the  $\Delta G_{\text{solv}}$  and the  $\Delta G_{\text{site}}$  were highly correlated

with each other and with the electrostatic interaction's energy components of both the site and solvent systems. Additionally, the VdW dispersion and VdW repulsion components were weakly inversely correlated. Finally, the  $\Delta\Delta G_{\text{binding}}$  was largely independent for the different energy components, although site VdW repulsion weakly correlated with  $\Delta\Delta G_{\text{binding}}$ , suggesting the importance of static interaction in determining the  $\Delta\Delta G_{\text{binding}}$  for the ERAG test set.

Table 7-2 – Discrimination between binders and decoys for  $\Delta\Delta G_{\text{binding}}$  energy components for the ERAG test set. Highlighted cell have a p-score less than 0.05.

|                                   |                            | Discrimination<br>T-test<br>(p-score) | Mean Energy<br>Binders<br>(kcal mol <sup>-1</sup> ) | Mean Energy<br>Decoys<br>(kcal mol <sup>-1</sup> ) |
|-----------------------------------|----------------------------|---------------------------------------|---|--|
| $\Delta\Delta G_{\text{binding}}$ |                            | 4.6E-14                               | -17.11 ± 1.60                                       | -9.39 ± 0.23                                       |
| Binding to Site                   | Free Energy                | 1.2E-02                               | -33.80 ± 2.00                                       | -31.08 ± 0.55                                      |
|                                   | Repulsion                  | 2.0E-05                               | 16.35 ± 1.27  | 19.35 ± 0.23                                       |
|                                   | Dispersion                 | 3.8E-03                               | -46.70 ± 1.06                                       | -45.07 ± 0.13                                      |
|                                   | Electrostatic              | 1.8E-01                               | -12.31 ± 1.87                                       | -13.67 ± 0.55                                      |
|                                   | Positional Constraints     | 7.8E-09                               | 13.55 ± 0.23  | 14.40 ± 0.14                                       |
|                                   | Conformational Constraints | 2.6E-29                               | 0.24 ± 0.03   | -0.05 ± 0.01                                       |
| Solvation                         | Free Energy                | 1.3E-04                               | -16.69 ± 2.35                                       | -21.66 ± 0.57                                      |
|                                   | Repulsion                  | 2.4E-02                               | 31.68 ± 0.51  | 32.29 ± 0.06                                       |
|                                   | Dispersion                 | 4.7E-01                               | -33.31 ± 0.52                                       | -33.51 ± 0.06                                      |
|                                   | Electrostatic              | 3.8E-05                               | -15.06 ± 2.34                                       | -20.43 ± 0.55                                      |

MF-FEB calculations were not able to discriminate between binders and decoys in the TRYP test set. While the mean  $\Delta\Delta G_{\text{binding}}$  for decoys was more reasonable for non-binding than ERAG decoys at -3 kcal mol<sup>-1</sup>, the mean  $\Delta\Delta G_{\text{binding}}$  for the binders was clearly incorrect at 1 kcal mol<sup>-1</sup>. Interestingly, both the  $\Delta G_{\text{solv}}$  and the  $\Delta G_{\text{site}}$  correctly discriminated between binders and decoys. In both cases, electrostatic interactions drove discrimination, consistent with the charged nature of the trypsin inhibitors. Additionally, the values of the energy component from electrostatic interactions were very low at less than -100 kcal mol<sup>-1</sup>. As with ERAG, we observed that the  $\Delta G_{\text{solv}}$  and the  $\Delta G_{\text{site}}$  were highly correlated with each other and with the electrostatic interaction's energy components and the VdW dispersion and repulsion components

were weakly inversely correlated. Finally, the  $\Delta\Delta G_{\text{binding}}$  for TRYP was also largely independent of the different energy components, although site positional constraints were weakly inversely correlated with  $\Delta\Delta G_{\text{binding}}$ , suggesting a reduction in entropy upon binding consistent with constraining the linear flexible trypsin binders.

Table 7-3 – Discrimination between binders and decoys for  $\Delta\Delta G_{\text{binding}}$  energy components for the TRYP test set. Highlighted cell have a p-score less than 0.05.

|                                   |                            | Discrimination<br>T-test<br>(p-score) | Mean Energy<br>Binders<br>(kcal mol <sup>-1</sup> ) | Mean Energy<br>Decoys<br>(kcal mol <sup>-1</sup> ) |
|-----------------------------------|----------------------------|---------------------------------------|---|--|
| $\Delta\Delta G_{\text{binding}}$ |                            | 1.2E-01                               | 1.16 ± 5.00   | -2.96 ± 0.74                                       |
| Binding to Site                   | Free Energy                | 6.1E-08                               | -108.02 ± 10.13                                     | -72.40 ± 2.10                                      |
|                                   | Repulsion                  | 8.8E-01                               | 31.08 ± 2.35  | 30.90 ± 0.36                                       |
|                                   | Dispersion                 | 5.4E-01                               | -50.74 ± 3.17                                       | -51.76 ± 0.46                                      |
|                                   | Electrostatic              | 3.1E-08                               | -95.02 ± 10.21                                      | -57.92 ± 2.04                                      |
|                                   | Positional Constraints     | 3.1E-01                               | 14.26 ± 0.96  | 14.84 ± 0.54                                       |
|                                   | Conformational Constraints | 5.1E-01                               | -0.76 ± 0.20  | -0.83 ± 0.03                                       |
| Solvation                         | Free Energy                | 1.1E-06                               | -109.37 ± 13.00                                     | -69.47 ± 2.25                                      |
|                                   | Repulsion                  | 4.0E-02                               | 43.94 ± 1.21  | 42.61 ± 0.13                                       |
|                                   | Dispersion                 | 1.8E-01                               | -49.00 ± 1.41                                       | -48.01 ± 0.16                                      |
|                                   | Electrostatic              | 4.6E-07                               | -102.84 ± 12.23                                     | -63.99 ± 2.22                                      |

Based on the strong correlation between  $\Delta G_{\text{solv}}$  and the  $\Delta G_{\text{site}}$  and the electrostatic interaction, we concluded that the first step in improving the MF-FEB calculations would be to improve the calculation of electrostatic interactions. This conclusion was additionally supported in the inability of MF-FEB calculations to discriminate between binders and decoys in the TYRP test set, while the  $\Delta G_{\text{solv}}$  and the  $\Delta G_{\text{site}}$  could discriminate. Because the energy from the electrostatic contributions was so larger, errors in their calculation would have a larger effect on the FEB, which was the difference between the two large values. An error of 10% would only vary the  $\Delta\Delta G_{\text{binding}}$  in the EARG test set by  $\sim 1$  kcal mol<sup>-1</sup>, while the same error would vary the  $\Delta\Delta G_{\text{binding}}$  in the TRYP test set by  $\sim 10$  kcal mol<sup>-1</sup> making a significant difference in

discriminating binders for non-binders. We discuss possible solutions to better model electrostatic interaction in Section 8.3.

Table 7-4 – Heat map of correlation between FEB energy components for the ERAG test set.

|                                   |                                   | Binding to Site |           |            |               |                        |                            | Solvation   |           |            |               |        |
|-----------------------------------|-----------------------------------|-----------------|-----------|------------|---------------|------------------------|----------------------------|-------------|-----------|------------|---------------|--------|
|                                   | $\Delta\Delta G_{\text{binding}}$ | Free Energy     | Repulsion | Dispersion | Electrostatic | Positional Constraints | Conformational Constraints | Free Energy | Repulsion | Dispersion | Electrostatic |        |
| $\Delta\Delta G_{\text{binding}}$ | 1.000                             | 0.179           | 0.475     | -0.099     | -0.028        | -0.124                 | -0.111                     | -0.249      | -0.138    | -0.015     | -0.235        |        |
| Binding to Site                   | Free Energy                       | 0.179           | 1.000     | 0.106      | -0.044        | 0.942                  | -0.042                     | -0.088      | 0.908     | 0.035      | 0.037         | 0.909  |
|                                   | Repulsion                         | 0.475           | 0.106     | 1.000      | -0.512        | -0.129                 | 0.269                      | -0.077      | -0.095    | 0.135      | -0.161        | -0.093 |
|                                   | Dispersion                        | -0.099          | -0.044    | -0.512     | 1.000         | -0.029                 | 0.178                      | -0.136      | -0.001    | -0.303     | 0.530         | -0.027 |
|                                   | Electrostatic                     | -0.028          | 0.942     | -0.129     | -0.029        | 1.000                  | 0.034                      | -0.057      | 0.940     | 0.059      | -0.020        | 0.944  |
|                                   | Positional Constraints            | -0.124          | -0.042    | 0.269      | 0.178         | 0.034                  | 1.000                      | -0.232      | 0.016     | 0.034      | 0.027         | 0.010  |
|                                   | Conformational Constraints        | -0.111          | -0.088    | -0.077     | -0.136        | -0.057                 | -0.232                     | 1.000       | -0.035    | -0.079     | 0.014         | -0.032 |
| Solvation                         | Free Energy                       | -0.249          | 0.908     | -0.095     | -0.001        | 0.940                  | 0.016                      | -0.035      | 1.000     | 0.095      | 0.044         | 0.994  |
|                                   | Repulsion                         | -0.138          | 0.035     | 0.135      | -0.303        | 0.059                  | 0.034                      | -0.079      | 0.095     | 1.000      | -0.540        | 0.045  |
|                                   | Dispersion                        | -0.015          | 0.037     | -0.161     | 0.530         | -0.020                 | 0.027                      | 0.014       | 0.044     | -0.540     | 1.000         | -0.008 |
|                                   | Electrostatic                     | -0.235          | 0.909     | -0.093     | -0.027        | 0.944                  | 0.010                      | -0.032      | 0.994     | 0.045      | -0.008        | 1.000  |

Table 7-5 – Heat map of correlation between FEB energy components for the TRYP test set.

|                                   |                                   | Binding to Site |           |            |               |                        |                            | Solvation   |           |            |               |        |
|-----------------------------------|-----------------------------------|-----------------|-----------|------------|---------------|------------------------|----------------------------|-------------|-----------|------------|---------------|--------|
|                                   | $\Delta\Delta G_{\text{binding}}$ | Free Energy     | Repulsion | Dispersion | Electrostatic | Positional Constraints | Conformational Constraints | Free Energy | Repulsion | Dispersion | Electrostatic |        |
| $\Delta\Delta G_{\text{binding}}$ | 1.000                             | -0.025          | -0.160    | 0.251      | -0.197        | -0.551                 | -0.044                     | -0.347      | -0.059    | -0.034     | -0.343        |        |
| Binding to Site                   | Free Energy                       | -0.025          | 1.000     | -0.232     | 0.255         | 0.968                  | -0.173                     | -0.117      | 0.946     | -0.098     | 0.147         | 0.946  |
|                                   | Repulsion                         | -0.160          | -0.232    | 1.000      | -0.736        | -0.239                 | 0.041                      | 0.052       | -0.171    | 0.122      | -0.156        | -0.165 |
|                                   | Dispersion                        | 0.251           | 0.255     | -0.736     | 1.000         | 0.214                  | 0.172                      | -0.250      | 0.167     | -0.198     | 0.285         | 0.156  |
|                                   | Electrostatic                     | -0.197          | 0.968     | -0.239     | 0.214         | 1.000                  | 0.034                      | -0.112      | 0.972     | -0.082     | 0.117         | 0.973  |
|                                   | Positional Constraints            | -0.551          | -0.173    | 0.041      | 0.172         | 0.034                  | 1.000                      | -0.183      | 0.015     | 0.005      | 0.001         | 0.015  |
|                                   | Conformational Constraints        | -0.044          | -0.117    | 0.052      | -0.250        | -0.112                 | -0.183                     | 1.000       | -0.098    | -0.043     | 0.004         | -0.096 |
| Solvation                         | Free Energy                       | -0.347          | 0.946     | -0.171     | 0.167         | 0.972                  | 0.015                      | -0.098      | 1.000     | -0.079     | 0.156         | 0.999  |
|                                   | Repulsion                         | -0.059          | -0.098    | 0.122      | -0.198        | -0.082                 | 0.005                      | -0.043      | -0.079    | 1.000      | -0.761        | -0.083 |
|                                   | Dispersion                        | -0.034          | 0.147     | -0.156     | 0.285         | 0.117                  | 0.001                      | 0.004       | 0.156     | -0.761     | 1.000         | 0.124  |
|                                   | Electrostatic                     | -0.343          | 0.946     | -0.165     | 0.156         | 0.973                  | 0.015                      | -0.096      | 0.999     | -0.083     | 0.124         | 1.000  |

## Chapter 8 Conclusions and Future Directions

Chapter 1 introduced computer-aided drug discovery, presented the current challenges inherent in virtual screening, and proposed combining virtual screening with MF-FEB rescoring as a method of overcoming those challenges. Chapter 2, Chapter 3, Chapter 5, and Chapter 6 described the implementation of MF-FEB rescoring and Chapters 4 and Chapter 7 described its testing in three systems. This chapter reviews the seven previous chapters, discussing their findings, how we could have been improved the work done, and what are the next steps in improving MF-FEB calculations and binding prediction.

This chapter is divided into four sections. Section 8.1 discusses the performance of MF-FEB rescoring within this work, and evaluates its utility as a drug discovery tool. Section 8.2 discusses retrospective improvements to this work. Section 8.3 discusses the next steps that one could take to improve MF-FEB rescoring. Section 8.4 concludes the chapter by discussing how this work informs the improvement of binding prediction.

### 8.1 MF-FEB RESCORING AS A TOOL FOR DRUG DISCOVERY

CADD tools can potentially lower the high cost of drug discovery thereby facilitating research for neglected diseases. (Trouiller, Olliaro et al. 2002; Nwaka and Ridley 2003; Stirner 2008) Therefore, the goal of this work was to improve CADD tools by improving the hit rate in docking-based virtual screening. As explained previously, current virtual screening tools, while enriching compound databases for potentially active compounds, generate large numbers of false positives leading to expensive and potentially fruitless bench-top testing. (Kellenberger, Rodrigo et al. 2004; Chen, Lyne et al. 2006; Warren, Andrews et al. 2006; Cross, Thompson et al. 2009) While docking programs are able to reproduce experimental binding poses, their scoring functions fail to reproduce accurately or correlate with experimental FEB, leading to poor enrichment in virtual screening. The assumption has been that the statistical mechanics perturbation based FEB calculations are more accurate than docking scoring functions, but their

computational intensity prevents their use in virtual screening. (Alonso, Bliznyuk et al. 2006; Gilson and Zhou 2007; Guimaraes and Cardozo 2008; Michel and Essex 2010) However, Roux and coworkers developed a mean field FEB method. It is amenable to a grid computing environment allowing perturbation based FEB calculations to be used in large-scale virtual screening projects. (Deng and Roux 2006; Wang, Deng et al. 2006; Deng and Roux 2008) Our hypothesis, therefore, was that the rescoring of docking generated poses with MF-FEB calculations would increase enrichment over the docking scoring function.

We tested our hypotheses using three test sets. In Chapter 4, we tested our hypothesis using the engineered hydrophobic binding pocket of L99A T4 lysozyme with a small compound database of 30 compounds. In Chapter 6, we tested our hypothesis using the ERAG and TRYP test sets from DUD, containing 2,428 and 1,587 compounds respectively. For all test sets, MF-FEB rescoring improved enrichment over AutoDock4's scoring function. However, MF-FEB rescoring was only able to discriminate between binders and non-binders, or decoys, in L99A T4 lysozyme and ERAG test sets. Additionally, MF-FEB rescoring did not outperform other scoring function in the ERAG and TRYP test sets.

We previously defined a good CADD tool as one that could identify bioactive compounds from a virtual chemical library, in a variety of different systems, and in a timely and efficient manner. So, does MF-FEB rescoring meet the four criteria of a good CADD tool? In both the ERAG and TRYP test set, MF-FEB rescoring had a higher success rate than AutoDock4's scoring function, even with the limitations of the MF-FEB calculations, suggesting that MF-FEB rescoring can improve the likelihood of identifying hits in a virtual screening project and thereby identify bioactive compounds. However, MF-FEB performance is system dependent as shown by comparing the enrichment of the ERAG and TRYP test sets. In addition, MF-FEB rescoring is currently not timely, in that the current runtime for a single target on World Community Grid is about six months, not including the time required to establish the project. Finally, MF-FEB rescoring is not efficient for two reasons. First, its current implementation

involves work unit processing steps that World Community Grid servers could perform, but are being performed on the project servers. Additionally, the FEB simulations use the general-purpose molecular modeling program CHARMM, which could be replaced with a program optimized to run MF-FEB calculations. Second, and most importantly, for both ERAG and TRYP there were less computationally demanding docking programs that outperformed MF-FEB rescoring in enriching compound databases.

## **8.2 IMPROVING DISCOVERING DENGUE DRUGS-TOGETHER**

The objective of this section is to discuss changes we would have made to this project by applying the knowledge gained during the course of the project. Knowledge was gained from the experiences of performing virtual screening with docking programs and FEB calculations, and from the concurrent advances in the field. We discuss how we would improve this project in two parts. In Section 8.2.1 we discuss what improvements we would like to have made to this project based on what would have been possible at the time we were doing the project. There are changes we would have liked to make to this project that were not possible when the project began, for example, the availability of the DUD test sets. Therefore, in Section 8.2.2 we outline what the best practices are for preparing to run a MF-FEB rescoring project using currently available tools.

### **8.2.1 Improvements to the Past**

There are a number of ways we could have improved the project. Potential improvements can be classified into three categories: first, are modifications to the test set, second, are enhancements to our testing of docking programs, and third, are modifications to MF-FEB calculations.

In the optimization of AutoDock4, we used only a small trypsin test set (see Section 3.1.1) and in the optimization of the MF-FEB calculations we used only the benzene-L99A T4 lysozyme system ( see Section 3.2.1). Ideally, we should have developed a unified and diverse

test set for docking and MF-FEB optimization that included both binders and non-binders. This ideal test set would have included the L99A T4 lysozyme test set and the trypsin test set, allowing for both the comparison to other work afforded by the L99A T4 lysozyme test set and the ligand and electrostatic complexity of the trypsin test set. Additionally, we would have added the HIV protease and neuraminidase test sets for the reasons discussed in Section 3.1.2.1. Given time and resources, it would have been ideal to measure the FEB of each compound, both binder and non-binders, in-house to provide a standardized test set. An accurate, unified, and diverse test set would allow for better optimization for AutoDock4 and the MF-FEB calculations, which would have supported or improved the parameters selected in Chapter 3. The improved test set would also have provided early indication of the poor performance of the TYRP test set.

Ideally, we would have performed the AutoDock4 virtual screening results filtering study described in Section 3.1.2 before the launch of Phase 1 of DDDT. Had we done those tests earlier, we would have taken the time to look at the performance of other docking programs in the virtual screening of trypsin, our dengue protease stand-in. Unfortunately, the DUD database was not established at the time we were preparing for Phase 1 and our computational resources were limited. We did virtually screen the NCI diversity set and the trypsin test set against 1eb2 trypsin structure resulting in a close to random distribution of binders, but did not yet understand the results in light of selecting ligands to progress from Phase 1 to Phase 2, and were depending on MF-FEB rescoring to improve enrichment.

We would like to have improved the initial testing of the MF-FEB calculations. However, the very steep learning curve and high computational demand of the MF-FEB calculations prevented us from doing tests then that we would otherwise choose to do now. We would have tested and optimized more parameters on a number of systems (at least eight different ligands, half binders and half non-binders) in four different targets as described above. Using these systems, we would have done the same work we did in Section 3.2.1 on run time and system size, but would have examined other parameters, like the constraint forces, and tested

different atomic partial charge models and force fields. Additionally, we should have spent additional time studying the variance in different MF-FEB calculations due to different random number seeds, as seen in Section 3.2.1, to better understand its cause. Finally, given time and resources, we would like to have tested MF-FEB rescoring on all the DUD test sets. A full DUD test would have provided more insights into how to improve MF-FEB calculations as different targets present different docking and rescoring challenges and may have even shown cases where MF-FEB does not improve enrichment.

Many of the modifications discussed above would have allowed us to identify the shortcomings in the MF-FEB calculations earlier in our work. However, we would have most likely spent time addressing those problems, or might have abandoned parts of the project, and might never have implemented Phase 2. Had we not done the MF-FEB calculations on World Community Grid, we would have been prevented from gaining the insights provided by rescoring the ERAG and TRYP test set.

### **8.2.2 Best Practices for Virtual Screening with MF-FEB Rescoring**

Based on the experience gained during this project, this section describes the best practices for performing a MF-FEB rescoring project. This section also focuses on the best practices for preparing for a virtual screening project, thereby ensuring the best use of MF-FEB rescoring.

Based on the results of this and other works, the docking programs and MF-FEB rescoring performances are target dependent. (Kellenberger, Rodrigo et al. 2004; Kontoyianni, McClellan et al. 2004; Warren, Andrews et al. 2006; Cross, Thompson et al. 2009) Therefore, before committing the computational resources required for MF-FEB rescoring, it is important to determine the potential performance of the system in a virtual screening project. Bench marking the system requires first obtaining or constructing a test set, second determining which docking

program or scoring function provides the best enrichment, and third determining how well the MF-FEB calculations can discriminate between binders and non-binders of the target.

The first step in testing a target system is to make a test set. The DUD test sets are examples of good test sets. The test set for the target system should be assembled containing binders and non-binders, or binders and decoys. Binders should be only a few percent of the total system, modeling the few binders in a virtual screening library. Ideally, the test set would come from experimental data on the target, for example a high throughput screening experiment. If that data were not available, data from related systems could be used, as we used trypsin as a stand-in for dengue protease. The test set should include at least one co-crystal structure of a ligand bound to the target, to test docking program's accuracy at pose prediction, and for benchmarking the accuracy of the MF-FEB calculations given an ideal docking pose. Additionally, a small subset, 4 to 10 compounds, of binder and non-binders should be assembled to parameterize and test the MF-FEB calculations.

In the second step, the test set should be virtually screened using different docking programs and scoring functions, determining which programs or function provides the best enrichment. For example, the full test set compiled in step one could be virtually screened using the freely available docking programs of DOCK (Ewing, Makino et al. 2001), AutoDock4 (Huey, Morris et al. 2007; Morris, Huey et al. 2009), and Vina (Trott and Olson 2010). The selection of which docking programs to test can be focused using data on related systems present in comparative docking studies like Warren et. al. (Warren, Andrews et al. 2006) and Cross et al. (Cross, Thompson et al. 2009). The docking programs should be tested using the same parameters and search volumes that will be used in the full virtual screening project. Using the docking program with the best enrichment will increase the probability that binders will be selected for rescoring as current grid computational resources prevent MF-FEB rescoring for more than a few thousand compounds.

The third and fourth steps is to optimize the MF-FEB calculations and then to determine if the MF-FEB calculations can discriminate between binders and non-binders of the target. This testing is done using the smaller test set compiled in the first step. The desired result is to see discrimination between binders and non-binders and have a difference in mean FEB comparable to the ERAG results to minimize overlap in the binder and non-binder populations, thereby reducing false positives. The MF-FEB parameters should be optimized for convergence. MF-FEB rescoring should not be used if convergence or effective discrimination cannot be achieved.

Based on the results of this work, the targets that are most likely to be successful in MF-FEB rescoring will be those that do not undergo significant conformational changes upon ligand binding and whose ligands are not charged.

### **8.3 FUTURE DIRECTIONS**

The goal of this work was to improve drug discovery tools by rescoring docking generated poses with MF-FEB calculations to improve virtual screening enrichment rates. However, while improving enrichment, MF-FEB rescoring was unable to discriminate effectively between binders and non-binders in the TRYP test set. (Chapter 7) This section proposes avenues of future research to address this limitation. The inability of the MF-FEB calculations to discriminate could arise from inaccuracies in the modeling of the system, parameter and force field selection, fundamental problems in the MF-FEB calculations, or any combination of the three. MF-FEB calculations do not model the energy contribution from conformation changes in the target that are not in the site system's explicit model or that do not occur within the simulation time scale. (See Chapter 3, Chapter 4) However, trypsin is stable receptor (Guvench, Price et al. 2005) that should have performed well in MF-FEB calculations. Therefore, target flexibility was likely not the problem. We will assume, for this section, that the fundamental theory behind the MF-FEB calculations was appropriate, which is supported by this work (see Chapter 1, Chapter 4, Chapter 7) and others (Deng and Roux 2006; Wang, Deng et al.

2006). Therefore, the problem of poor discrimination in the TRYP test set likely resulted from the parameterization or force field selection.

As discussed in Chapter 7, the  $\Delta G_{\text{site}}$  and  $\Delta G_{\text{solv}}$  absolute values were significantly larger in the TRYP test set than in the ERAG test set, while their mean difference,  $\Delta\Delta G_{\text{binding}}$ , remained within an experimentally appropriate range. Those larger energies in the TRYP test set correlated with large energy contributions from the charge-charge interactions, which suggested that the accuracy of the trypsin binding calculations would be largely influenced by the accuracy of the charge-charge interaction component. (See Table 7-2, Table 7-3, Table 7-4, and Table 7-5) The dependence on electrostatic interactions for binding was consistent with the structure of trypsin and its inhibitors; this has also been demonstrated in other work. (Shi, Jiao et al. 2009) On average,  $\Delta\Delta G_{\text{binding}}$  values were only few percent of the total  $\Delta G_{\text{site}}$  and  $\Delta G_{\text{solv}}$  values. Therefore, a small error the calculations of  $\Delta G_{\text{site}}$  or  $\Delta G_{\text{solv}}$  would be sufficient to change the binding status of ligand. Since  $\Delta G_{\text{site}}$  and  $\Delta G_{\text{solv}}$  were largely composed of the charge-charge interaction in the TRYP MF-FEB calculations, improvement in the accuracy of their calculations could improve MF-FEB's ability to discriminate between binders and non-binders in the TRYP. The charge-charge interactions for the FEB were determined by the decoupling of those interactions as described by the CHARMM force field using Coulomb's Law and the partial charges assigned to the atoms. Therefore, to improve these calculations could require modifying the decoupling steps to increase sampling, modifying the partial charges, and/or changing the force field.

The following sections explore improving MF-FEB's treatment of charge-charge interactions. Section 8.3.1 describes the tests sets for analyzing and improving the MF-FEB calculations. Section 8.3.2 discusses modifications to the MF-FEB calculations that will be tested. Section 8.3.3 discusses the potential results and conclusions of those tests.

### 8.3.1 Test systems

In order to improve the treatment of charge-charge interactions in MF-FEB calculations, test sets should be compiled that can examine the accuracy of the FEB prediction for trypsin binders, test discrimination between trypsin binders and non-binders, and ensure the MF-FEB calculations retain their performance established in other systems. The test sets should be small as the MF-FEB calculations are computationally intensive and the calculations will need to be performed multiple times under different conditions. The test set should use experimentally determined structures, because they represent the ideal docking result and, therefore, remove possible errors arising from the ligand pose in the MF-FEB calculations. (See Chapter 4) The test set would be composed of twelve protein:ligand systems from three groups; trypsin binders, trypsin non-binders, and the control systems.

The first group would contain four trypsin binders covering a range of FEB at  $\sim 2$  kcal mol<sup>-1</sup> steps and include benzamidine. The  $\sim 2$  kcal mol<sup>-1</sup> step is greater than the average error for MF-FEB calculations, as determined in section 4.2, allowing the FEB predictions to not only be checked for accuracy, but to be checked for accurate ordering of binders. Benzamidine provides a binding standard for comparison with the non-binders discussed below. The systems in this group include 1G36 ( $\Delta G_{\text{exp}} = -10.2$  kcal mol<sup>-1</sup> (Nar, Bauer et al. 2001)), 1EB2 ( $\Delta G_{\text{exp}} = -8.2$  kcal mol<sup>-1</sup> (Liebeschuetz, Jones et al. 2002)), 1BTY ( $\Delta G_{\text{exp}} = -7.2$  to  $-6.8$  kcal mol<sup>-1</sup> (Katz, Finer-Moore et al. 1995)), and 1UTN ( $\Delta G_{\text{exp}} = -4.7$  kcal mol<sup>-1</sup> (Leiros, Brandsdal et al. 2004)). All these ligands are basic, consistent with the chemistry of trypsin inhibitors.

The second group would contain four systems of two compounds that did not bind to trypsin each with two poses. These systems determine if modifications to the MF-FEB calculations would improve discrimination between binders and non-binders and examine the modeling of the

charge-charge interactions. The non-binders would be isopropylbenzene and benzoic acid.<sup>19</sup> It is expected that in the docked structures neither the propyl group of isopropylbenzene nor the carboxylic group of benzoic acid will overlay the diamine of benzamidine, as the P1 pocket is negatively charged. Therefore, two poses will be used for each non-binding ligand. The first pose will be the docking-generated pose, to examine the FEB predictions based on the interaction suggested from docking and mimicking virtual screening results. The second pose would come from modifying the benzamidine in 1BTY to the two ligands to examine the modeling of the charge-charge interaction between the functional groups of the ligands and the P1 site.<sup>20</sup>

The last group of four systems would be a control set to insure that modifications made to the MF-FEB calculations were not biasing the FEB predictions to discriminate trypsin binders from non-binders, but were improving the accuracy of the MF-FEB prediction. This group would contain two binding and two docked non-binder systems from the L99A T4 lysozyme test set. The binders would be benzene ( $\Delta G_{\text{exp}} = -5.2 \text{ kcal mol}^{-1}$  (See Section 4.2)) and ethylbenzene ( $\Delta G_{\text{exp}} = -5.8 \text{ kcal mol}^{-1}$ , 1NHB (Morton, Baase et al. 1995; Morton and Matthews 1995) . (See Chapter 4) The benzene system allows for comparison to previous optimization work. (See Chapter 3) Ethylbenzene serves not only as a positive control for binding to L99A T4 lysozyme, but also allows for limited comparison with the trypsin non-binder controls. The group's non-

---

<sup>19</sup> The non-binding compounds could have been obtained the literature, but most trypsin non-binders for that source were complex molecules developed for thrombin or factor X inhibition. Therefore, it was decided to propose ligands that based on their chemistry should not bind, yet would enable focused inquires into the charge-charge interaction. However, their non-binding status should be determined experimentally.

<sup>20</sup> The positional constraints in the MF-FEB calculations should the set to the engineered poses, and not the equilibrium pose to insure interactions between the functional group and the P1 pocket.

binders should be benzamidine and benzoic acid, consistent with the trypsin binder and non-binder controls.

### 8.3.2 Methods

This section discusses potential modifications to the MF-FEB calculations to improve discrimination in the TRYP test set. The general approach would be to test each modification independently, using the test set described in section 8.3.1, and comparing the MF-FEB predicted energies to the experimental FEB for each system. As the test set is composed of 12 systems, simulations can be performed using the high-performance computing resources at TACC as described in Section 2.1.2. Modifications to the MF-FEB calculations would include increasing the number of decoupling steps to improve sampling, modifying the partial charges, and changing the force field.

In the current implementation of the MF-FEB calculations, the  $\lambda$  steps for decoupling the charge-charge interaction were 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1. (See Chapter 2) These steps provided sufficient sampling for convergence of the WHAM analysis for most MF-FEB calculations. (See Section 6.2.3.2 for a discussion on MF-FEB failure rates.) These steps were selected from work with the L99A T4 lysozyme system, which did not have a charged binding site. (Deng and Roux 2006) In the L99A T4 lysozyme system, plots of charge-charge interaction energy vs.  $\lambda$  step were linear in the site calculation for benzene and indole, but non-linear in the bulk calculation for benzene and indole, and non-linear in the site and bulk calculations for phenol. (See Figures 6, 7, and 8 in (Deng and Roux 2006)) As interactions between the ligand and the target were coupled, the charge-charge interaction energies changed as if the charges were moved together. The charge-charge interactions were modeled using Coulomb's Law; therefore, the energies increased as lambda approached one and stopped at the energy based on the physical distance between the charges. In trypsin, a plot of the charge-charge interaction energy vs.  $\lambda$  step would be expected to be non-linear, similar to curves for the bulk and polar ligand site calculations from the L99A T4 lysozyme system. It is possible that

when large charge-charge interactions were present, the transition for  $\lambda$  between 0.9 and 1 was under-sampled in our current algorithm. To test if sampling frequency was causing errors in the charge-charge interaction energy of MF-FEB simulations, we would perform the MF-FEB calculations adding additional  $\lambda$  steps in the 0.9 to 1 range distributed using a  $1/x$  relationship.

Different partial charge assignment methods have been tested with the MF-FEB calculations and showed only small differences in FEB predictions. However, these partial charge comparisons have only been reported for the benzene-L99A T4 lysozyme system. (Deng and Roux 2006) Therefore, different partial charge models should be tested in the trypsin system using the test set described in section 8.3.1 to evaluate the impact of partial charge models on FEB predictions. There are several of methods for assigning partial charge that can be tested, from the empirical to the quantum mechanical. The testing of different semiempirical partial charge models for the ligands would begin with those available in Antechamber (Wang, Wang et al. 2006) and include the TPACM4 charge set, which was specifically developed for the GAFF (Mukherjee, Patra et al. 2011). Attention will be paid to the partial charges on benzamidine diamine, ensuring that the charges were distributed according to the resonance structure and that the positive charge was distributed between the amine nitrogens and the shared adjacent cationic carbon. Computationally more demanding ab-initio quantum mechanical methods for assigning partial changes should be tested (e.g. SCF/6-31G\*, MP2/cc-pVTZ) as opposed to the quicker empirical and elector population based methods; however, recent studies showed little differences between free energy of solvation calculations using quantum mechanical and other partial change assignment methods. (Mobley, Dumont et al. 2007; Zhou, Huang et al. 2010)

As discussed in Section 3.2, different force fields were tested with the MF-FEB simulation and produced  $\Delta\Delta G_{\text{binding}}$  predictions that were within the error of the calculations; however, these studies only examined the lysozyme system. (Wang, Deng et al. 2006) Therefore, a systemic test of different force fields using the test set described in section 8.3.1 would allow for effect of the force field on MF-FEB simulations to be determined. Initially, the

AMBER force field (Cornell, Cieplak et al. 1995) from which the GAFF force field was derived would be examined. (Wang, Wolf et al. 2004) Next polarizable force fields (PFF) would be tested to reduce errors in the charge-charge interactions in the TRYP test set. The PFF models the dipole moment and polarizability of atom by extending the force field's non-bonding interaction terms. (Halgren and Damm 2001; Lopes, Roux et al. 2009) PFFs have been used to determine the free energy of solvation for methane, ethane and methanol using free energy perturbation methods, alchemy and annihilation, with an error of 0.13 kcal mol<sup>-1</sup>. (Kaminski, Ponomarev et al. 2009) PFFs have been used to improve the FEB prediction of benzamidine to trypsin, but not using MF-FEB calculations. (Jiao, Golubkov et al. 2008) CHARMM already has two polarizable force fields (Anisimov, Vorobyov et al. 2004; Patel and Brooks 2004; Patel, Mackerell et al. 2004) that could be used for testing PFF with MF-FEB calculations; however, ligand parameterization may be a challenge as ligand parameters are not included in the standard force field and both PFFs are under development (Lopes, Roux et al. 2009). Regardless of implementation, using a PFF will likely increase the runtime of the MD simulations especially when employing a fluctuating charge model PFF. (Lopes, Roux et al. 2009)

### **8.3.3 Potential Results and Conclusions**

The challenge of modeling charge-charge interactions is that the absolute energetic contribution to FEB from charge-charge interactions for a system cannot be determined experimentally. Therefore, success in the modification of charge-charge interactions in MF-FEB calculations can only be judged on the modification's ability to "better" reproduce the experimentally determined FEB and not on their ability to model the charge-charge interactions. Ultimately, making these modifications is akin to changing the weight of empirical parameters, thereby reducing these modifications potential effectiveness in other systems.

The ideal result would be to have a modification to the charge-charge model of the MF-FEB calculations that would accurately predict the FEB for each binder in the test set and predict the non-binders FEB to be  $\sim 0$  kcal mol<sup>-1</sup> or greater. A good result would contain small

errors in the accuracy of the FEB prediction, but would correctly rank-order the trypsin binders, and discriminate the binders and non-binders with the minimum mean separation of  $\sim 7 \text{ kcal mol}^{-1}$  (i.e., the difference between non-binding and the mean FEB of the trypsin binders). The minimally acceptable result would be to correctly discriminate binders from non-binders. An ideal result would indicate an improvement in the MF-FEB calculations, which could then be tested with a larger test set to determine the robustness of the modification. However, good or acceptable results would indicate the MF-FEB calculations needed further improvement in order to have the predicted FEB match the experimental FEB. A failure to obtain improvement would suggest that a combination of modifications should be tested, that other parameters (e.g. the constraint model) should be tested, or that there were fundamental problems in the MF-FEB approach and/or the force fields.

The above suggestions to improve the MF-FEB calculations (i.e., modifying the decoupling step size, modifying the partial charges, and changing the force field) were presented in order of implementation difficulty. However, the most difficult implementation, PFFs, may provide the best improvement to MF-FEB calculations. Increasing the  $\lambda$  step size may improve the accuracy of the charge-charge interaction, but the energy curve was not asymptotic as  $\lambda$  approached one. Therefore, the energy between  $\lambda$  steps 0.9 and 1 as currently implemented may be sufficient. Changes in partial charge assignment should only cause scaling of the energy determined from Coulomb's Law for any pairwise interaction. Therefore, unless there were significant changes in the distribution of the partial charges for different models, it may be difficult for charge models to improve the demarcation between binders and non-binders. It is most likely that the force field will need to include changes in polarization as charged ligands are brought near a charged binding site. PFFs have been used to improve the FEB prediction of benzamidine to trypsin (Jiao, Golubkov et al. 2008), and it has been shown that the FEB correlates with the polarizability of the ligand. (Shi, Jiao et al. 2009) Therefore, it is most likely

that PFF will improve the ability of MF-FEB calculations to discriminate binders from non-binders in charged systems.

#### **8.4 IMPROVING DRUG DISCOVERY TOOLS**

Ultimately, it is our ability to accurately predict solubility and binding that will allow us to effectively use computational methods for drug discovery. The goal of this section is to examine the impact of this work on the field of virtual screening and FEB predictions, and suggest what directions the field may take in improve FEB predictions.

The findings of this work are significant to the fields of virtual screening and FEB predictions. The MF-FEB rescoring of the ERAG and TRYP represent the largest implementation of perturbation free energy binding calculations so far. Previous studies have been small (< 60 compounds) and largely methods orientated (Deng and Roux 2009; Michel and Essex 2010), however this work used two orders of magnitude larger test sets that could be compared to docking scoring functions.

The field recognizes the challenges of accurate FEB prediction both in the error in the predictions and in the limitations of the force fields (Merz 2010; Michel and Essex 2010), and the difficulty of modeling solvation and entropy in docking scoring functions. Nevertheless, it has been a common assumption that rigorous perturbation based FEB calculations are “better” than empirical scoring functions used by many docking programs due to their sampling of microstates to determine the FEB. (Alonso, Bliznyuk et al. 2006; Gilson and Zhou 2007; Guimaraes and Cardozo 2008; Michel and Essex 2010) However, this work has shown that this assumption is incorrect in the case of MF-FEB rescoring, and that perturbation based FEB calculations perform similarly to empirical scoring functions in that they produce false positives and perform differently in different systems. The continual presences of a large number false positives in any virtual screening project using any scoring method suggests a fundamental problem in our

computational models, be it in the description of the molecules (e.g., force fields), or in our approach to modeling binding both mathematically and in simulation.

We currently are unable to accurately and robustly predict the absolute FEB using empirical methods, be it scoring functions, LIE or force fields, because all empirical methods are limited by their test set. Therefore, until first principles methods can be used, all methods will fail when they are used for a target they cannot model. Research is being done to include QM modeling in CADD (Zhou, Huang et al. 2010), but, like SM approaches to FEB prediction, our ability to implement it is limited by the computational power available to the researchers.

Nevertheless, we can improve our approach to empirical parameterization and the development and test of SM-FEB methods by including non-binders in our training set. While a few non-binders have been used in the testing of perturbation based FEB methods (Deng and Roux 2009; Michel and Essex 2010), most docking scoring functions do not. It should be no surprise when a mean score for a virtual screening project suggests that most compounds in the compound database are binders, when the scoring function is parameterized to find binders. Therefore, the next improvement in the field of drug discovery may not be a better method of determined bindings, but a more effective method of determining non-binders, because a few positive results with a number of false negatives is more useful to drug discovery then trying to find a few true positives hidden in among many false positives.

## Bibliography

Screensave Lifesaver.

- Alam, M. A. and P. K. Naik (2009). "Applying linear interaction energy method for binding affinity calculations of podophyllotoxin analogues with tubulin using continuum solvent model and prediction of cytotoxic activity." J Mol Graph Model **27**(8): 930-43.
- Allen, F. H. (2002). "The Cambridge Structural Database: a quarter of a million crystal structures and rising." Acta Crystallogr B **58**(Pt 3 Pt 1): 380-8.
- Alonso, H., A. A. Bliznyuk, et al. (2006). "Combining docking and molecular dynamic simulations in drug design." Med Res Rev **26**(5): 531-68.
- Anderson, A. C. (2003). "The process of structure-based drug design." Chem Biol **10**(9): 787-97.
- Andricopulo, A. D., L. B. Salum, et al. (2009). "Structure-based drug design strategies in medicinal chemistry." Curr Top Med Chem **9**(9): 771-90.
- Anisimov, V. M., I. V. Vorobyov, et al. (2004). "CHARMM all-atom polarizable force field parameter development for nucleic acids." Biophysical Journal **86**(1): 415a-415a.
- Aqvist, J. and J. Marelus (2001). "The linear interaction energy method for predicting ligand binding free energies." Comb Chem High Throughput Screen **4**(8): 613-26.
- Aqvist, J., C. Medina, et al. (1994). "A new method for predicting binding affinity in computer-aided drug design." Protein Eng **7**(3): 385-91.
- Aronovitz, L. G. (2006). *New Drug Development: Science, Business, Regulatory, and Intellectual Property Issues Cited as Hampering Drug Development Efforts*. U. S. G. A. Office.
- Barril, X., P. Brough, et al. (2005). "Structure-based discovery of a new class of Hsp90 inhibitors." Bioorg Med Chem Lett **15**(23): 5187-91.
- Baxter, A. D. and P. M. Lockey (2001). "'Hit' to 'lead' and 'lead' to 'candidate' optimisation using multi-parametric principles." Drug Discovery World **Winter 2001/2**.
- Bhagal, N. and M. Balls (2008). "Translation of new technologies: from basic research to drug discovery and development." Curr Drug Discov Technol **5**(3): 250-62.
- Bissantz, C., P. Bernard, et al. (2003). "Protein-based virtual screening of chemical databases. II. Are homology models of G-Protein Coupled Receptors suitable targets?" Proteins **50**(1): 5-25.
- Bleicher, K. H., H. J. Bohm, et al. (2003). "Hit and lead generation: beyond high-throughput screening." Nat Rev Drug Discov **2**(5): 369-78.
- Bohm, H. J. (1994). "The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure." J Comput Aided Mol Des **8**(3): 243-56.
- Bohm, H. J. (1998). "Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs." J Comput Aided Mol Des **12**(4): 309-23.
- Boresch, S., F. Tettinger, et al. (2003). "Absolute binding free energies: A quantitative approach for their calculation." Journal of Physical Chemistry B **107**(35): 9535-9551.
- Brooks, B. R., R. E. Bruccoleri, et al. (1983). "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations." Journal of Computational Chemistry **4**: 187-217.

- Brown, S. P. and S. W. Muchmore (2006). "High-throughput calculation of protein-ligand binding affinities: modification and adaptation of the MM-PBSA protocol to enterprise grid computing." *J Chem Inf Model* **46**(3): 999-1005.
- Bullard, D., A. Gobbi, et al. (2008). "Hydra: a self regenerating high performance computing grid for drug discovery." *J Chem Inf Model* **48**(4): 811-6.
- Bursulaya, B. D., M. Totrov, et al. (2003). "Comparative study of several algorithms for flexible ligand docking." *J Comput Aided Mol Des* **17**(11): 755-63.
- Campbell, S. J., N. D. Gold, et al. (2003). "Ligand binding: functional site location, similarity and docking." *Curr Opin Struct Biol* **13**(3): 389-95.
- Carlsson, J., L. Boukharta, et al. (2008). "Combining Docking, Molecular Dynamics and the Linear Interaction Energy Method to Predict Binding Modes and Affinities for Non-nucleoside Inhibitors to HIV-1 Reverse Transcriptase." *J Med Chem*.
- Chang, M. W., W. Lindstrom, et al. (2007). "Analysis of HIV wild-type and mutant structures via in silico docking against diverse ligand libraries." *J Chem Inf Model* **47**(3): 1258-62.
- Charifson, P. S., J. J. Corkery, et al. (1999). "Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins." *J Med Chem* **42**(25): 5100-9.
- Chen, H., P. D. Lyne, et al. (2006). "On evaluating molecular-docking methods for pose prediction and enrichment factors." *J Chem Inf Model* **46**(1): 401-15.
- Chen, H. M., B. F. Liu, et al. (2007). "SODOCK: swarm optimization for highly flexible protein-ligand docking." *J Comput Chem* **28**(2): 612-23.
- Chen, J., M. Yang, et al. (2009). "Insights into the functional role of protonation states in the HIV-1 protease-BEA369 complex: molecular dynamics simulations and free energy calculations." *J Mol Model* **15**(10): 1245-52.
- Chen, J., S. Zhang, et al. (2009). "Insights into drug resistance of mutations D30N and I50V to HIV-1 protease inhibitor TMC-114: free energy calculation and molecular dynamic simulation." *J Mol Model* **16**(3): 459-68.
- Clark, M., F. Guarnieri, et al. (2006). "Grand canonical Monte Carlo simulation of ligand-protein binding." *J Chem Inf Model* **46**(1): 231-42.
- Clark, M., S. Meshkat, et al. (2009). "Grand canonical free-energy calculations of protein-ligand binding." *J Chem Inf Model* **49**(4): 934-43.
- Clark, R. D., A. Strizhev, et al. (2002). "Consensus scoring for ligand/protein interactions." *J Mol Graph Model* **20**(4): 281-95.
- Collins, M. D., G. Hummer, et al. (2005). "Cooperative water filling of a nonpolar protein cavity observed by high-pressure crystallography and simulation." *Proc Natl Acad Sci U S A* **102**(46): 16668-71.
- Cornell, W. D., P. Cieplak, et al. (1995). "A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules." *Journal of the American Chemical Society* **117**(19): 5179-5197.
- Cossins, B. P., S. Foucher, et al. (2009). "Assessment of nonequilibrium free energy methods." *J Phys Chem B* **113**(16): 5508-19.
- Cross, J. B., D. C. Thompson, et al. (2009). "Comparison of several molecular docking programs: pose prediction and virtual screening accuracy." *J Chem Inf Model* **49**(6): 1455-74.
- Cross, J. B., D. C. Thompson, et al. (2009). "Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy." *J Chem Inf Model*.

- Cummings, M. D., R. L. DesJarlais, et al. (2005). "Comparison of automated docking programs as virtual screening tools." J Med Chem **48**(4): 962-76.
- Das, D., Y. Koh, et al. (2009). "Prediction of potency of protease inhibitors using free energy simulations with polarizable quantum mechanics-based ligand charges and a hybrid water model." J Chem Inf Model **49**(12): 2851-62.
- Davis, I. W. and D. Baker (2009). "RosettaLigand docking with full ligand and receptor flexibility." J Mol Biol **385**(2): 381-92.
- De Azevedo, W. F., Jr. (2010). "MolDock applied to structure-based virtual screening." Curr Drug Targets **11**(3): 327-34.
- Degliesposti, G., V. Kasam, et al. (2009). "Design and discovery of plasmepsin II inhibitors using an automated workflow on large-scale grids." ChemMedChem **4**(7): 1164-73.
- DeLano, W. L. (2008). The PyMOL Molecular Graphics System. Palo Alto, CA, USA., DeLano Scientific LLC.
- Deng, Y. and B. Roux (2006). "Calculation of Standard Binding Free Energies: Aromatic Molecules in T4 Lysozyme L99A Mutant." Journal of Chemical Theory and Computation **2**(5): 1225-1273.
- Deng, Y. and B. Roux (2008). "Computation of binding free energy with molecular dynamics and grand canonical Monte Carlo simulations." J Chem Phys **128**(11): 115103.
- Deng, Y. and B. Roux (2009). "Computations of standard binding free energies with molecular dynamics simulations." J Phys Chem B **113**(8): 2234-46.
- Deprez-Poulain, R. and B. Deprez (2004). "Facts, figures and trends in lead generation." Curr Top Med Chem **4**(6): 569-80.
- Drews, J. (2000). "Drug discovery: a historical perspective." Science **287**(5460): 1960-4.
- Duff, M. R., V. K. Mudhivarthi, et al. (2009). "Rational design of anthracene-based DNA binders." J Phys Chem B **113**(6): 1710-21.
- Eldridge, M. D., C. W. Murray, et al. (1997). "Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes." J Comput Aided Mol Des **11**(5): 425-45.
- Engel, T. (2006). "Basic overview of chemoinformatics." J Chem Inf Model **46**(6): 2267-77.
- Erbel, P., N. Schiering, et al. (2006). "Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus." Nat Struct Mol Biol **13**(4): 372-3.
- Eriksson, A. E., W. A. Baase, et al. (1992). "A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene." Nature **355**(6358): 371-3.
- Ewing, T. J., S. Makino, et al. (2001). "DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases." J Comput Aided Mol Des **15**(5): 411-28.
- Federsel, H. J. (2006). "In search of sustainability: process R&D in light of current pharmaceutical industry challenges." Drug Discov Today **11**(21-22): 966-74.
- Fischer, M. and R. E. Hubbard (2009). "Fragment-based ligand discovery." Mol Interv **9**(1): 22-30.
- Freceer, V. and S. Miertus (2010). "Design, structure-based focusing and in silico screening of combinatorial library of peptidomimetic inhibitors of Dengue virus NS2B-NS3 protease." J Comput Aided Mol Des **24**(3): 195-212.
- Friesner, R. A., J. L. Banks, et al. (2004). "Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy." J Med Chem **47**(7): 1739-49.
- Gamblin, S. J. and J. J. Skehel (2010). "Influenza hemagglutinin and neuraminidase membrane glycoproteins." J Biol Chem **285**(37): 28403-9.

- Gershell, L. J. and J. H. Atkins (2003). "A brief history of novel drug discovery technologies." Nat Rev Drug Discov **2**(4): 321-7.
- Ghosh, S., A. Nie, et al. (2006). "Structure-based virtual screening of chemical libraries for drug discovery." Current Opinion in Chemical Biology **10**(3): 194-202.
- Gilson, M. K., J. A. Given, et al. (1997). "The statistical-thermodynamic basis for computation of binding affinities: a critical review." Biophys J **72**(3): 1047-69.
- Gilson, M. K. and H. X. Zhou (2007). "Calculation of protein-ligand binding affinities." Annu Rev Biophys Biomol Struct **36**: 21-42.
- Goodsell, D. S. and A. J. Olson (1990). "Automated docking of substrates to proteins by simulated annealing." Proteins **8**(3): 195-202.
- Gorelik, B. and A. Goldblum (2008). "High quality binding modes in docking ligands to proteins." Proteins **71**(3): 1373-86.
- Grant, M. A. (2009). "Protein structure prediction in structure-based ligand design and virtual screening." Comb Chem High Throughput Screen **12**(10): 940-60.
- Guimaraes, C. R. and M. Cardozo (2008). "MM-GB/SA rescoring of docking poses in structure-based lead optimization." J Chem Inf Model **48**(5): 958-70.
- Guimaraes, C. R. and A. M. Mathiowetz (2010). "Addressing limitations with the MM-GB/SA scoring procedure using the WaterMap method and free energy perturbation calculations." J Chem Inf Model **50**(4): 547-59.
- Guvench, O. and A. D. MacKerell, Jr. (2008). "Comparison of protein force fields for molecular dynamics simulations." Methods Mol Biol **443**: 63-88.
- Guvench, O. and A. D. MacKerell, Jr. (2009). "Computational fragment-based binding site identification by ligand competitive saturation." PLoS Comput Biol **5**(7): e1000435.
- Guvench, O., D. J. Price, et al. (2005). "Receptor rigidity and ligand mobility in trypsin-ligand complexes." Proteins **58**(2): 407-17.
- Halgren, T. A. and W. Damm (2001). "Polarizable force fields." Curr Opin Struct Biol **11**(2): 236-42.
- Halgren, T. A., R. B. Murphy, et al. (2004). "Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening." J Med Chem **47**(7): 1750-9.
- Halperin, I., B. Ma, et al. (2002). "Principles of docking: An overview of search algorithms and a guide to scoring functions." Proteins **47**(4): 409-43.
- Hartmann, C., I. Antes, et al. (2009). "Docking and scoring with alternative side-chain conformations." Proteins **74**(3): 712-26.
- Helms, V. and R. Wade (1998). "Computational alchemy to calculate absolute protein-ligand binding free energy." J. Am. Chem. Soc. **120**: 2710-2713.
- Hodgson, J. (2001). "ADMET--turning chemicals into drugs." Nat Biotechnol **19**(8): 722-6.
- Huang, N., B. K. Shoichet, et al. (2006). "Benchmarking sets for molecular docking." J Med Chem **49**(23): 6789-801.
- Huey, R., G. M. Morris, et al. (2007). "A semiempirical free energy force field with charge-based desolvation." Journal of Computational Chemistry **28**(6): 1145-1152.
- Irwin, J. J. and B. K. Shoichet (2005). "ZINC--a free database of commercially available compounds for virtual screening." J Chem Inf Model **45**(1): 177-82.
- Jacq, N., J. Salzemann, et al. (2006). "Demonstration of in silico docking at a large scale on grid infrastructure." Stud Health Technol Inform **120**: 155-7.
- Jacq, N. B., Vincent (2007). "Virtual screening on large scale grids." Parallel Computing **33**: 289-301.

- Jakalian, A., D. B. Jack, et al. (2002). "Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation." Journal of Computational Chemistry **23**(16): 1623-1641.
- Jiao, D., P. A. Golubkov, et al. (2008). "Calculation of protein-ligand binding free energy by using a polarizable potential." Proc Natl Acad Sci U S A **105**(17): 6290-5.
- Jiao, D., J. Zhang, et al. (2009). "Trypsin-ligand binding free energies from explicit and implicit solvent simulations with polarizable potential." J Comput Chem **30**(11): 1701-11.
- Jones, G., P. Willett, et al. (1997). "Development and validation of a genetic algorithm for flexible docking." J Mol Biol **267**(3): 727-48.
- Jorgensen, W. L. (2004). "The many roles of computation in drug discovery." Science **303**(5665): 1813-8.
- Jorgensen, W. L. (2009). "Efficient drug lead discovery and optimization." Acc Chem Res **42**(6): 724-33.
- Jorgensen, W. L., J. K. Buckner, et al. (1988). "Efficient Computation of Absolute Free-Energies of Binding by Computer-Simulations - Application to the Methane Dimer in Water." Journal of Chemical Physics **89**(6): 3742-3746.
- Jorgensen, W. L. and J. Tiradorives (1988). "The Opls Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin." Journal of the American Chemical Society **110**(6): 1657-1666.
- Kaminski, G. A., S. Y. Ponomarev, et al. (2009). "Polarizable Simulations with Second order Interaction Model - force field and software for fast polarizable calculations: Parameters for small model systems and free energy calculations." J Chem Theory Comput **5**(11): 2935-2943.
- Kasam, V., J. Salzemann, et al. (2009). "WISDOM-II: Screening against multiple targets implicated in malaria using computational grid infrastructures." Malar J **8**(1): 88.
- Kasam, V., M. Zimmermann, et al. (2007). "Design of new plasmepsin inhibitors: a virtual high throughput screening approach on the EGEE grid." J Chem Inf Model **47**(5): 1818-28.
- Katz, B. A., J. Finer-Moore, et al. (1995). "Episelection: novel Ki approximately nanomolar inhibitors of serine proteases selected by binding or chemistry on an enzyme surface." Biochemistry **34**(26): 8264-80.
- Kellenberger, E., J. Rodrigo, et al. (2004). "Comparative evaluation of eight docking tools for docking and virtual screening accuracy." Proteins **57**(2): 225-42.
- Keseru, G. M. and G. M. Makara (2006). "Hit discovery and hit-to-lead approaches." Drug Discov Today **11**(15-16): 741-8.
- Kitano, H. (2002). "Systems biology: a brief overview." Science **295**(5560): 1662-4.
- Kitchen, D. B., H. Decornez, et al. (2004). "Docking and scoring in virtual screening for drug discovery: methods and applications." Nat Rev Drug Discov **3**(11): 935-49.
- Knox, A. J., M. J. Meegan, et al. (2005). "Considerations in compound database preparation--"hidden" impact on virtual screening results." J Chem Inf Model **45**(6): 1908-19.
- Kollman, P. A., I. Massova, et al. (2000). "Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models." Acc Chem Res **33**(12): 889-97.
- Kontoyianni, M., L. M. McClellan, et al. (2004). "Evaluation of docking performance: comparative data on docking algorithms." J Med Chem **47**(3): 558-65.
- Kontoyianni, M., G. S. Sokol, et al. (2005). "Evaluation of library ranking efficacy in virtual screening." J Comput Chem **26**(1): 11-22.

- Kortagere, S. and S. Ekins (2010). "Troubleshooting computational methods in drug discovery." J Pharmacol Toxicol Methods **61**(2): 67-75.
- Kosugi, T., I. Nakanishi, et al. (2009). "Binding free energy calculations of adenosine deaminase inhibitor and the effect of methyl substitution in inhibitors." J Chem Inf Model **49**(3): 615-22.
- Kramer, B., M. Rarey, et al. (1999). "Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking." Proteins **37**(2): 228-41.
- Kumar, S., J. M. Rosenberg, et al. (1992). "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method." Journal of Computational Chemistry **13**(8): 1011-1021.
- Kuntz, I. D., J. M. Blaney, et al. (1982). "A geometric approach to macromolecule-ligand interactions." J Mol Biol **161**(2): 269-88.
- Leach, A. R., B. K. Shoichet, et al. (2006). "Prediction of protein-ligand interactions. Docking and scoring: successes and gaps." J Med Chem **49**(20): 5851-5.
- Leiros, H. K., B. O. Brandsdal, et al. (2004). "Trypsin specificity as elucidated by LIE calculations, X-ray structures, and association constant measurements." Protein Sci **13**(4): 1056-70.
- Li, H., H. Zhang, et al. (2009). "An effective docking strategy for virtual screening based on multi-objective optimization algorithm." BMC Bioinformatics **10**: 58.
- Liebeschuetz, J. W., S. D. Jones, et al. (2002). "PRO\_SELECT: combining structure-based drug design and array-based chemistry for rapid lead discovery. 2. The development of a series of highly potent and selective factor Xa inhibitors." J Med Chem **45**(6): 1221-32.
- Lipinski, C. A., F. Lombardo, et al. (2001). "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings." Adv Drug Deliv Rev **46**(1-3): 3-26.
- Lipsky, M. S. and L. K. Sharp (2001). "From idea to market: the drug approval process." J Am Board Fam Pract **14**(5): 362-7.
- Lopes, P. E., B. Roux, et al. (2009). "Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability. Theory and applications." Theor Chem Acc **124**(1-2): 11-28.
- Macarron, R. (2006). "Critical review of the role of HTS in drug discovery." Drug Discov Today **11**(7-8): 277-9.
- MacKerell, A. D., D. Bashford, et al. (1998). "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins." J. Phys. Chem. B **102**(18): 3586-3616.
- Mackerell, A. D., Jr. (2004). "Empirical force fields for biological macromolecules: overview and issues." J Comput Chem **25**(13): 1584-604.
- MacKerell Jr, A. D., C. L. Brooks III, et al. (1998). CHARMM: The Energy Function and Its Parameterization with an Overview of the Program, John Wiley & Sons: Chichester. **1**: 271-277.
- Mahoney, M. W. and W. L. Jorgensen (2000). "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions." J. Chem. Phys. **112**: 8910-8922.
- Meiler, J. and D. Baker (2006). "ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility." Proteins **65**(3): 538-48.
- Merz, K. M. (2010). "Limits of Free Energy Computation for Protein-Ligand Interactions." J Chem Theory Comput **6**(4): 1018-1027.

- Michel, J. and J. W. Essex (2010). "Prediction of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations." J Comput Aided Mol Des.
- Michielan, L. and S. Moro (2010). "Pharmaceutical perspectives of nonlinear QSAR strategies." J Chem Inf Model **50**(6): 961-78.
- Mitscher, L. A. and A. Dutta (2006). Contemporary Drug Discovery. Drug Discovery and Development. M. S. Chorghade. Hoboken, John Wiley & Sons, Inc. **1**: 103-128.
- Mizutani, M. Y. and A. Itai (2004). "Efficient method for high-throughput virtual screening based on flexible docking: discovery of novel acetylcholinesterase inhibitors." J Med Chem **47**(20): 4818-28.
- Mizutani, M. Y., Y. Takamatsu, et al. (2006). "Effective handling of induced-fit motion in flexible docking." Proteins **63**(4): 878-91.
- Mobley, D. L., J. D. Chodera, et al. (2006). "On the use of orientational restraints and symmetry corrections in alchemical free energy calculations." J Chem Phys **125**(8): 084902.
- Mobley, D. L., J. D. Chodera, et al. (2007). "The Confine-and-Release Method: Obtaining Correct Binding Free Energies in the Presence of Protein Conformational Change." J Chem Theory Comput **3**(4): 1231-1235.
- Mobley, D. L. and K. A. Dill (2009). "Binding of small-molecule ligands to proteins: "what you see" is not always "what you get"." Structure **17**(4): 489-98.
- Mobley, D. L., E. Dumont, et al. (2007). "Comparison of charge models for fixed-charge force fields: small-molecule hydration free energies in explicit solvent." J Phys Chem B **111**(9): 2242-54.
- Mobley, D. L., A. P. Graves, et al. (2007). "Predicting absolute ligand binding free energies to a simple model site." J Mol Biol **371**(4): 1118-34.
- Mohan, V., A. C. Gibbs, et al. (2005). "Docking: successes and challenges." Curr Pharm Des **11**(3): 323-33.
- Morens, D. M. and A. S. Fauci (2008). "Dengue and hemorrhagic fever: a potential threat to public health in the United States." JAMA **299**(2): 214-6.
- Morris, G. M., D. S. Goodsell, et al. (1998). "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function." Journal of Computational Chemistry **19**(14): 1639-1662.
- Morris, G. M., R. Huey, et al. (2009). "Autodock4 and AutoDockTools4: Automated docking with selective receptor flexibility." J Comput Chem.
- Morton, A., W. A. Baase, et al. (1995). "Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme." Biochemistry **34**(27): 8564-75.
- Morton, A. and B. W. Matthews (1995). "Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: linkage of dynamics and structural plasticity." Biochemistry **34**(27): 8576-88.
- Mrazek, M. F. and E. Mossialos (2003). "Stimulating pharmaceutical research and development for neglected diseases." Health Policy **64**(1): 75-88.
- Mukherjee, G., N. Patra, et al. (2011). "A fast empirical GAFF compatible partial atomic charge assignment scheme for modeling interactions of small molecules with biomolecular targets." J Comput Chem **32**(5): 893-907.
- Murray, C. W., C. A. Baxter, et al. (1999). "The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase." J Comput Aided Mol Des **13**(6): 547-62.

- Naggie, S., K. Patel, et al. (2010). "Hepatitis C virus directly acting antivirals: current developments with NS3/4A HCV serine protease inhibitors." J Antimicrob Chemother **65**(10): 2063-9.
- Nar, H., M. Bauer, et al. (2001). "Structural basis for inhibition promiscuity of dual specific thrombin and factor Xa blood coagulation inhibitors." Structure **9**(1): 29-37.
- Nervall, M., P. Hanspers, et al. (2008). "Predicting Binding Modes from Free Energy Calculations." J Med Chem.
- Noble, C. G., Y. L. Chen, et al. (2010). "Strategies for development of Dengue virus inhibitors." Antiviral Res **85**(3): 450-62.
- Nwaka, S. and R. G. Ridley (2003). "Virtual drug discovery and development for neglected diseases through public-private partnerships." Nat Rev Drug Discov **2**(11): 919-28.
- Oprea, T. I., T. K. Allu, et al. (2007). "Lead-like, drug-like or "Pub-like": how different are they?" J Comput Aided Mol Des **21**(1-3): 113-9.
- Oprea, T. I., A. M. Davis, et al. (2001). "Is there a difference between leads and drugs? A historical perspective." J Chem Inf Comput Sci **41**(5): 1308-15.
- Pang, Y. P., E. Perola, et al. (2001). "EUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases." J Comput Chem **22**(15): 1750-1771.
- Park, J. W. and W. H. Jo (2010). "Computational design of novel, high-affinity neuraminidase inhibitors for H5N1 avian influenza virus." Eur J Med Chem **45**(2): 536-41.
- Patel, S. and C. L. Brooks, 3rd (2004). "CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations." J Comput Chem **25**(1): 1-15.
- Patel, S., A. D. Mackerell, Jr., et al. (2004). "CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model." J Comput Chem **25**(12): 1504-14.
- Paul, N. and D. Rognan (2002). "ConsDock: A new program for the consensus analysis of protein-ligand interactions." Proteins **47**(4): 521-33.
- Paul, S. M., D. S. Mytelka, et al. "How to improve R&D productivity: the pharmaceutical industry's grand challenge." Nat Rev Drug Discov **9**(3): 203-14.
- Perdih, A., U. Bren, et al. (2009). "Binding free energy calculations of N-sulphonyl-glutamic acid inhibitors of MurD ligase." J Mol Model **15**(8): 983-96.
- Ponder, J. W. and D. A. Case (2003). "Force fields for protein simulations." Adv Protein Chem **66**: 27-85.
- Rarey, M., B. Kramer, et al. (1996). "A fast flexible docking method using an incremental construction algorithm." J Mol Biol **261**(3): 470-89.
- Rastelli, G., A. Del Rio, et al. (2010). "Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA." J Comput Chem **31**(4): 797-810.
- Richards, W. G. (2002). "Virtual screening using grid computing: the screensaver project." Nat Rev Drug Discov **1**(7): 551-5.
- Salzemann, J., V. Kasam, et al. (2007). "Grid enabled high throughput virtual screening against four different targets implicated in malaria." Stud Health Technol Inform **126**: 47-54.
- Schrattenholz, A. and V. Soskic (2008). "What does systems biology mean for drug development?" Curr Med Chem **15**(15): 1520-8.
- Service, R. F. (2004). "Surviving the blockbuster syndrome." Science **303**(5665): 1796-9.

- Shi, Y., D. Jiao, et al. (2009). "Trypsin-ligand binding free energy calculation with AMOEBA." Conf Proc IEEE Eng Med Biol Soc **2009**: 2328-31.
- Shoichet, B. K. (2004). "Virtual screening of chemical libraries." Nature **432**(7019): 862-5.
- Singh, N. and A. Warshel (2010). "A comprehensive examination of the contributions to the binding entropy of protein-ligand complexes." Proteins **78**(7): 1724-35.
- Smith, R. H., Jr., W. L. Jorgensen, et al. (1998). "Prediction of binding affinities for TIBO inhibitors of HIV-1 reverse transcriptase using Monte Carlo simulations in a linear response method." J Med Chem **41**(26): 5272-86.
- Solis, F. J. and R. J. B. Wets (1981). "Minimization by Random Search Techniques." Mathematics of Operations Research **6**(1): 19-30.
- Stirner, B. (2008). "Stimulating research and development of pharmaceutical products for neglected diseases." Eur J Health Law **15**(4): 391-409.
- Stouch, T. R., J. R. Kenyon, et al. (2003). "In silico ADME/Tox: why models fail." J Comput Aided Mol Des **17**(2-4): 83-92.
- Takeda-Shitaka, M., D. Takaya, et al. (2004). "Protein structure prediction in structure based drug design." Curr Med Chem **11**(5): 551-8.
- Talele, T. T., S. A. Khedkar, et al. (2010). "Successful applications of computer aided drug discovery: moving drugs from concept to the clinic." Curr Top Med Chem **10**(1): 127-41.
- Teague, S. J., A. M. Davis, et al. (1999). "The Design of Leadlike Combinatorial Libraries." Angew Chem Int Ed Engl **38**(24): 3743-3748.
- Teodoro, M. L. and L. E. Kavraki (2003). "Conformational flexibility models for the receptor in structure based drug design." Curr Pharm Des **9**(20): 1635-48.
- Tetko, I. V., P. Bruneau, et al. (2006). "Can we estimate the accuracy of ADME-Tox predictions?" Drug Discov Today **11**(15-16): 700-7.
- Thompson, D. C., C. Humblet, et al. (2008). "Investigation of MM-PBSA rescoring of docking poses." J Chem Inf Model **48**(5): 1081-91.
- Thomsen, R. and M. H. Christensen (2006). "MolDock: a new technique for high-accuracy molecular docking." J Med Chem **49**(11): 3315-21.
- Tomlinson, S. M., R. D. Malmstrom, et al. (2009). "Structure-based discovery of dengue virus protease inhibitors." Antiviral Res **82**(3): 110-4.
- Tomlinson, S. M., R. D. Malmstrom, et al. (2009). "New approaches to structure-based discovery of dengue protease inhibitors." Infect Disord Drug Targets **9**(3): 327-43.
- Tomlinson, S. M. and S. J. Watowich (2008). "Substrate inhibition kinetic model for West Nile virus NS2B-NS3 protease." Biochemistry **47**(45): 11763-70.
- Tomlinson, S. M. and S. J. Watowich (2011). "Anthracene-based inhibitors of dengue virus NS2B-NS3 protease." Antiviral Res **89**(2): 127-35.
- Totrov, M. and R. Abagyan (2008). "Flexible ligand docking to multiple receptor conformations: a practical alternative." Curr Opin Struct Biol **18**(2): 178-84.
- Trott, O. and A. J. Olson (2010). "Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading." Journal of Computational Chemistry **31**(2): 455-461.
- Trouiller, P., P. Olliaro, et al. (2002). "Drug development for neglected diseases: a deficient market and a public-health policy failure." Lancet **359**(9324): 2188-94.
- Verdonk, M. L., V. Berdini, et al. (2004). "Virtual screening using protein-ligand docking: avoiding artificial enrichment." J Chem Inf Comput Sci **44**(3): 793-806.

- Verdonk, M. L., G. Chessari, et al. (2005). "Modeling water molecules in protein-ligand docking using GOLD." *J Med Chem* **48**(20): 6504-15.
- Verdonk, M. L., J. C. Cole, et al. (2003). "Improved protein-ligand docking using GOLD." *Proteins* **52**(4): 609-23.
- Villoutreix, B. O., R. Eudes, et al. (2009). "Structure-based virtual ligand screening: recent success stories." *Comb Chem High Throughput Screen* **12**(10): 1000-16.
- Wang, J., Y. Deng, et al. (2006). "Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials." *Biophys J* **91**(8): 2798-814.
- Wang, J., W. Wang, et al. (2006). "Automatic atom type and bond type perception in molecular mechanical calculations." *J Mol Graph Model* **25**(2): 247-60.
- Wang, J., R. M. Wolf, et al. (2004). "Development and testing of a general amber force field." *J Comput Chem* **25**(9): 1157-74.
- Wang, R., L. Lai, et al. (2002). "Further development and validation of empirical scoring functions for structure-based binding affinity prediction." *J Comput Aided Mol Des* **16**(1): 11-26.
- Wang, W., O. Donini, et al. (2001). "Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions." *Annu Rev Biophys Biomol Struct* **30**: 211-43.
- Wang, W., J. Wang, et al. (1999). "What determines the van der Waals coefficient beta in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations?" *Proteins* **34**(3): 395-402.
- Warren, G. L., C. W. Andrews, et al. (2006). "A critical assessment of docking programs and scoring functions." *J Med Chem* **49**(20): 5912-31.
- Weeks, J. D., D. Chandler, et al. (1971). "Role of repulsive forces in determining equilibrium structure of simple liquids." *J. Chem. Phys.* **54**: 5237-5247.
- Weininger, D. (1988). "Smiles, a Chemical Language and Information-System .1. Introduction to Methodology and Encoding Rules." *Journal of chemical information and computer sciences* **28**(1): 31-36.
- Weininger, D. (1990). "Smiles .3. Depict - Graphical Depiction of Chemical Structures." *Journal of chemical information and computer sciences* **30**(3): 237-243.
- Weininger, D., A. Weininger, et al. (1989). "Smiles .2. Algorithm for Generation of Unique Smiles Notation." *Journal of chemical information and computer sciences* **29**(2): 97-101.
- Woo, H. J. and B. Roux (2005). "Calculation of absolute protein-ligand binding free energy from computer simulations." *Proc Natl Acad Sci U S A* **102**(19): 6825-30.
- You, L. (2004). "Toward computational systems biology." *Cell Biochem Biophys* **40**(2): 167-84.
- Yusof, R., S. Clum, et al. (2000). "Purified NS2B/NS3 serine protease of dengue virus type 2 exhibits cofactor NS2B dependence for cleavage of substrates with dibasic amino acids in vitro." *J Biol Chem* **275**(14): 9963-9.
- Zhang, S. (2011). "Computer-aided drug discovery and development." *Methods Mol Biol* **716**: 23-38.
- Zhou, J. Z. (2011). "Cheminformatics and library design." *Methods Mol Biol* **685**: 27-52.
- Zhou, T., D. Huang, et al. (2010). "Quantum mechanical methods for drug design." *Curr Top Med Chem* **10**(1): 33-45.

## Vita

Robert Dean Malmstrom was born the 15th of May 1977 in Santa Clara California to Richard Dean Malmstrom and Anick Elaine Bertrand Malmstrom. He grew up in Utah, Minnesota, and Texas where he excelled in his academic work despite being dyslexic. After serving a mission for his church in Portugal, he pursued his interest in chemistry by obtaining a B.S. in biochemistry from Brigham Young University. During his undergraduate studies Robert began his research career in the lab of Dr. Gary Booth, first with insect behavioral studies, and then natural products research, leading Robert to a deep interest in drug discovery which he pursued into his graduate studies. Robert attended graduate school at the University of Texas Medical Branch working in the lab of Dr. Stanly Watowich. His graduate work focused on improving computational drug discovery methods by coupling perturbation based free energy of binding calculations with current virtual screening methods; collaborating with IBM's World Community Grid to launch Discovering Dengue Drugs-Together; and Influenza Antiviral Drug Search drug discovery projects. In his personal life, Robert is a husband and father, active in his church, and enjoys a variety of hobbies. He married his college sweetheart Elise Marie Hammer. They are the parents of two boys.

### Education

B.S., April 2002, Brigham Young University, Provo, UT

### Publications

Robert D. Malmstrom and Stanly J. Watowich; *Using Free Energy of Binding Calculations to Improve the Accuracy of Virtual Screening Predictions*, Journal of Chemical Information and Modeling, 2011 Jun 22. PMID: 21696204

Andrew Russo, Robert D. Malmstrom, Mark White, and Stanly J. Watowich; *Structural basis for substrate specificity of alphavirus nsP2 proteases*, Journal of Molecular Graphics and Modeling, 2010 Aug 24, 29(1):46-53. PMID: 20483643

Suzanne Tomlinson\*, Robert D. Malmstrom\*, and Stanly J. Watowich; *New approaches to structure-based discovery of dengue protease inhibitors*, Infectious Disorders Drug Targets, Jun. 2009, 9(3):327-43 PMID: 1919486

Suzanne Tomlinson, Robert D. Malmstrom, Andrew Russo, Nike Muller, Y.-P. Pang, and Stanly J. Watowich; *Structure-based discovery of dengue virus protease inhibitors*, Antiviral Research, Jun. 2008, 82(3):110-4 PMID: 19428601

## Summary of Dissertation

Current docking programs can accurately reproduce experimentally determined binding poses of small molecules to macromolecular targets. However, the empirical based scoring functions of docking programs do not accurately predict the free energy of binding, leading to large numbers of false positives when docking programs are used for virtual screening. To improve virtual screening enrichment, mean field perturbation based free energy of binding calculations were coupled with a traditional docking program to rescore docking generated poses. The rescoring method was tested in the L99A T4 lysozyme system and the Database of Useful Decoy's trypsin and estrogen receptor agonists test sets. IBM's World Community Grid was employed due to the computational demands of the free energy of binding calculations launching the projects Discovering Dengue Drug-Together and Influenza Antiviral Drug Search. Although rescoring docking generated poses with free energy of binding calculations consistently improved enrichment in all test sets, rescoring did not out perform all scoring functions.