

**UCSF**

**UC San Francisco Electronic Theses and Dissertations**

**Title**

Probing the Boundaries of Molecular Docking with Decoys and Model Systems

**Permalink**

<https://escholarship.org/uc/item/0v607948>

**Author**

Graves, Alan P

**Publication Date**

2007-09-27

Peer reviewed|Thesis/dissertation

Probing the Boundaries of Molecular Docking with Decoys and Model Systems

by

Alan P. Graves, III

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biophysics

in the

UMI Number: 3289301

Copyright 2007 by  
Graves, Alan P., III

All rights reserved.

UMI<sup>®</sup>

---

UMI Microform 3289301

Copyright 2008 by ProQuest Information and Learning Company.  
All rights reserved. This microform edition is protected against  
unauthorized copying under Title 17, United States Code.

---

ProQuest Information and Learning Company  
300 North Zeeb Road  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

Copyright 2007  
by  
Alan P. Graves, III

I would like to dedicate this dissertation to  
my parents, family, friends, and mentors.

## Acknowledgements

The successful completion of a graduate career is due largely in part to the support of an invaluable group of people. I feel fortunate to have had such a support system in my life. An essence of all those who I have come in contact have made me the person I am today. First, I would like to thank those people who have been most instrumental in my achievements at Northwestern University and especially UCSF. I thank Brian Shoichet, my advisor, for accepting me into his lab with open arms and providing me with the tools to follow my goals. He believed in me even when I did not believe in myself, and he has pushed me to reach heights that I did not imagine possible. I also thank the many members of the Shoichet Lab past and present for their scientific guidance and friendship: Federica Morandi (my rotation mentor and first friend in the lab), Binqing Wei, Ruth Brenk, Sarah Boyce, Andrea McReynolds, and Niu Huang (who have all helped immensely with the model binding sites either through docking, experiment, or both), Veena Thomas (who is smart *and* intelligent), Viva Habbit Van Assen, Yu Chen, Sandri Soelaiman, Brian Feng, Kerim Babaoglu, Kristin Coan, Rafaela Ferrieria (who was one of my rotation students), John Irwin, Kaushik Raha (who I will have the pleasure of working with at GSK), Johannes Hermann, Jerome Hert, Mike Keiser, Michael Mysinger, Pacal Wassam and others.

I would also like to thank the friends and collaborators that have made this work possible: David Mobley, Devleena Shivakumar, John Chodera, and David Case. I also thank my friends Alex So, Ray Nagatani, and Holly Atkinson. Special thanks go to Julie Ransom and the Biophysics Graduate Group who took me in as one of their own during my second year.

I am grateful for the members of my thesis committee, Ken Dill and Matt Jacobson. They are in many ways responsible for my successes; not to mention that Matt was a collaborator on the work discussed in Chapter 2, and Ken, who initially suggested

the idea for decoys, was a collaborator on the work discussed in Chapter 3. They often provided a fresh insight and angle—or as Ken would say “the view from 30,000 feet in the air”—that one often needs when so consumed with the nitty-gritty details of one’s own project.

I would especially like to thank Joshua Middleton whose love and support have carried me through the last three years of my graduate career. I am particularly grateful for his support over the last few months while I have been finishing up at UCSF. I would not have made it without him. I am so thankful for his warm heart, strength, positive outlook, love of life, understanding, creativity and faith in me.

Finally I would like to thank my family. My parents, Pete and Carolyn Graves, have been a constant source of love and support every step of the way. Ever since that morning of July 30<sup>th</sup>, 1979, they have been there for me through thick and thin. Without them none of this would have been possible. I thank my brothers Tom and Rodney Floyd for putting up with me when I was their annoying kid brother and for their love.

This dissertation is mainly composed of my two first-author publications and a third paper on which I shared first-authorship with David Mobley. Each chapter is prefaced by a gloss in which I have written the major points and motivations. I conclude this oeuvre with some thoughts of future directions.

The material in Chapter 1 first appeared in the *Journal of Medicinal Chemistry* in 2005. It was reproduced here with the permission of Graves, A.P.; Brenk, R.; Shoichet, B.K. Decoys for Docking. *J. Med. Chem.* **2005**, 48, 3714-3728. © 2005 *American Chemical Society*. The supplementary material from this publication has been included here as Appendix A.

The material in Chapter 2 will be submitted to the *Journal of Molecular Biology* in September 2007. It was reproduced here with the permission of Graves, A.P.;

Shivakumar, D.M.; Boyce, S.E.; Jacobson, M.P.; Case, D.A.; Shoichet, B.K. Rescoring Docking Hit Lists for Model Cavity Sites: Predictions and Experimental testing. *J. Mol. Biol.* **2007**. © 2007 Elsevier. The supplementary material from this publication has been included here as Appendix B.

The material in Chapter 3 first appeared in the *Journal of Molecular Biology* in 2007. It was reproduced here with the permission of Mobley, D.L.; Graves, A.P.; Chodera, J.D.; McReynolds, A.C.; Shoichet, B.K.; Dill, K.A. Predicting Absolute Ligand Binding Free Energies to a Simple Model Site. *J. Mol. Biol.* **2007**, 371, 1118-1134. © 2007 Elsevier.



## Abstract

### Probing the Boundaries of Molecular Docking with Decoys and Model Systems

Alan P. Graves, III

There are a plethora of computational tools available in the field of structure-based drug design that attempt to predict the complementarity and binding energetics of a ligand to a drug target of interest. These computational tools often comprise large errors and predict incorrect ligand geometries or select non-binding molecules over true ligands. While these false-positive hits, which are referred to as “decoys”, are frustrating, they potentially provide important tests for computational methods. In order to test the utility of these computational tools as well as suggest areas for improvement, I provide a database of decoys from several model binding sites which range in complexity from small cavities to real drug targets. Especially in the cavity sites, which are very simple, these decoys highlight particular weaknesses in sampling and scoring procedures.

With decoys from model binding sites, I examine molecular docking methods, which are used to screen large compound databases to find drug leads. Many of the decoy molecules that make up the high failure rate of docking screens are informative, arguably more so than successful predictions. Second, I examine MM-GBSA rescoring of docking hit lists. These more physically realistic methods have improved models for solvation, electrostatics, and conformational change than do most docking programs. MM-GBSA rescoring with binding site minimization improved the separation of known ligands from known decoys for each of the cavity sites and rescued docking false negatives, but also introduced several new decoys into the top-ranking molecules. Finally, I examine alchemical free energy calculations to predict accurate binding free energies of ligands to

the simplest hydrophobic cavity site. This method computed absolute binding free energies with an RMS of 1.9 kcal/mol for a set of known ligands, correctly discriminated between several true ligands and decoys in a set of putative binders, and calculated binding free energies of these with an RMS error of just 0.6 kcal/mol. I consider the origins of the successes and some of the particular sources of failures in docking and rescoring in the model sites as well as the implications for biologically relevant targets.

## TABLE OF CONTENTS

<b>List of Tables .....</b>	<b>xi</b>
<b>List of Figures.....</b>	<b>xiv</b>
<b>Introduction.....</b>	<b>1</b>
<b>Gloss to Chapter 1.....</b>	<b>6</b>
<b>1. Decoys for Docking .....</b>	<b>9</b>
Abstract.....	10
Introduction.....	12
Results.....	16
Discussion.....	38
Materials and Methods.....	42
Acknowledgements.....	48
<b>Gloss to Chapter 2.....</b>	<b>49</b>
<b>2. Rescoring Docking Hit Lists for Model Cavity Sites: Predictions and Experimental Testing.....</b>	<b>51</b>
Abstract.....	52
Introduction.....	55
Results.....	61
Discussion.....	84
Materials and Methods.....	91
Acknowledgments.....	96
<b>Gloss to Chapter 3.....</b>	<b>97</b>
<b>3. Predicting Absolute Ligand Binding Free Energies to a Simple Model Site.....</b>	<b>99</b>
Abstract.....	100
Introduction.....	102
Results.....	106

Discussion.....	134
Materials and Methods.....	138
Acknowledgments.....	149
<b>Future Directions .....</b>	<b>150</b>
<b>References.....</b>	<b>155</b>
<b>Appendix A. Supporting Material for “Decoys for Docking” .....</b>	<b>173</b>
<b>Appendix B. Supporting Material for “Rescoring Docking Hit Lists for Model Cavity Sites: Predictions and Experimental Testing” .....</b>	<b>190</b>

## List of Tables

### Chapter 1.

<b>Table 1.</b> Characteristic geometric decoys and native-like dockings assessed by different scoring functions.....	18
<b>Table 2.</b> Characteristic L99A experimentally tested ligands scoring in the top 10,000 docking hits and their ranks by different scoring functions.....	23
<b>Table 3.</b> Characteristic L99A experimentally tested decoys scoring in the top 10,000 docking hits and their ranks by different scoring functions.....	24
<b>Table 4.</b> Characteristic L99A/M102Q experimentally tested ligands and decoys scoring in the top 10,000 docking hits and their ranks by different scoring functions. .	26
<b>Table 5.</b> Decoys for L99A and L99A/M102Q predicted by the Screenscore, FlexX, PLP, PMF, and SMOG scoring functions. ....	31
<b>Table 6.</b> Characteristic AmpC ligands and decoys and their ranks by different scoring functions.....	34
<b>Table 7.</b> Decoys for AmpC predicted by the Screenscore, FlexX, PLP, PMF, and SMOG scoring functions.....	36

### Chapter 2.

<b>Table 1.</b> Compounds predicted by AMBERDOCK and PLOP to bind to L99A.....	67
<b>Table 2.</b> Crystallographic data and predicted ligand RMSD's for L99A and L99A/M102Q.....	70
<b>Table 3.</b> Compounds predicted by AMBERDOCK and PLOP to bind to L99A/M102Q. ....	72
<b>Table 4.</b> Compounds predicted to bind by AMBERDOCK and PLOP to CCP. ....	76

<b>Table 5.</b> Crystallographic data and predicted ligand RMSD's for CCP. ....	79
<b>Table 6.</b> Likely ligands and decoys among the top 100 ranked ligands by docking and MM-GBSA. ....	82

### Chapter 3.

<b>Table 1.</b> Calculated and experimental binding free energies for ligands of the apolar binding site.....	107
<b>Table 2.</b> Binding free energies calculated for selected ligands using their <i>holo</i> structures as a starting point. ....	115
<b>Table 3.</b> Calculated and experimental binding free energies for ligands of the apolar binding site using AM1-BCC charges. ....	118
<b>Table 4.</b> Novel compounds for which predictions were made and later tested experimentally. ....	125
<b>Table 5.</b> X-ray data collection and refinement. ....	128

### Appendix A.

<b>Table S1.</b> Ligand bound protein structures from the PDB used to test for geometric decoys.....	174
<b>Table S2.</b> Experimentally tested L99A Ligands. ....	175
<b>Table S3.</b> Experimentally tested L99A/M102Q ligands. ....	176
<b>Table S4.</b> Experimentally tested L99A decoys. ....	177
<b>Table S5.</b> Experimentally tested L99A/M102Q decoys.....	178
<b>Table S6.</b> Experimentally tested AmpC ligands and decoys. ....	179
<b>Table S7.</b> Crystal data for catechol in complex with L99A/M102Q. ....	181

## Appendix B.

<b>Table S1.</b> Top 100 hits predicted by DOCK3.5.54 for L99A.....	195
<b>Table S2.</b> Top 100 hits predicted by PLOP for L99A.....	197
<b>Table S3.</b> Top 100 hits predicted by AMBERDOCK for L99A.....	199
<b>Table S4.</b> Top 100 hits predicted by DOCK3.5.54 for L99A/M102Q.....	201
<b>Table S5.</b> Top 100 hits predicted by PLOP for L99A/M102Q.....	203
<b>Table S6.</b> Top 100 hits predicted by AMBERDOCK for L99A/M102Q.....	205
<b>Table S7.</b> Top 100 hits predicted by DOCK3.5.54 for CCP.....	207
<b>Table S8.</b> Top 100 hits predicted by PLOP for CCP.....	209
<b>Table S9.</b> Top 100 hits predicted by AMBERDOCK for CCP.....	211

## List of Figures

### Chapter 1.

<b>Figure 1.</b> Protein targets used for hit list decoys.....	21
<b>Figure 2.</b> Catechol bound to L99A/M102Q.....	27
<b>Figure 3.</b> Enrichment of ligands for L99A, L99A/M102Q, and AmpC .....	29
<b>Figure 4.</b> Characteristic high scoring docking hits to L99A by six scoring functions.....	32

### Chapter 2.

<b>Figure 1.</b> The model cavity sites.....	59
<b>Figure 2.</b> Retrospective enrichment of ligands and decoys for cavity sites.....	63
<b>Figure 3.</b> Predicted and experimental ligand orientations for the hydrophobic cavity....	69
<b>Figure 4.</b> Predicted and experimental ligand orientations for the polar cavity.....	74
<b>Figure 5.</b> Predicted and experimental ligand orientations for the anionic cavity. ....	78
<b>Figure 6.</b> Topologically similar ligands and decoys.....	88

### Chapter 3.

<b>Figure 1.</b> The model hydrophobic binding site in the L99A mutant of T4 lysozyme. ..	104
<b>Figure 2.</b> Calculated binding free energies compared with experiment. ....	109
<b>Figure 3.</b> Val111 reorients on ligand binding.....	112
<b>Figure 4.</b> Dock scores versus experimental binding free energies.....	120
<b>Figure 5.</b> Comparison of calculated and experimental binding free energies with the protein held rigid. ....	122
<b>Figure 6.</b> Five compounds for which binding predictions were made.....	126



<b>Figure 7.</b> Predicted and experimental ligand orientations.....	130
<b>Figure 8.</b> Representative ITC data. ....	133

### **Appendix A.**

<b>Figure S1.</b> Score versus RMSD from native pose for ligand ALG in thrombin using seven scoring functions.....	182
<b>Figure S2.</b> Score versus RMSD bounding curves for native-like and geometric decoys. ....	183
<b>Figure S3.</b> Minimum bounding curves of DOCK score vs. RMSD from crystallographic pose for DHFR ligands.....	185
<b>Figure S4.</b> Minimum bounding curves of DOCK score vs. RMSD from crystallographic pose for thrombin ligands.....	186
<b>Figure S5.</b> Minimum bounding curves of DOCK score vs. RMSD from crystallographic pose for TS ligands.....	187
<b>Figure S6.</b> Minimum bounding curves of DOCK score vs. RMSD from crystallographic pose for PNP ligands.....	188
<b>Figure S7.</b> Minimum bounding curves of DOCK score vs. RMSD from crystallographic pose for AChE ligands. ....	189

### **Appendix B.**

<b>Figure S1.</b> Flowchart and parameters for AMBERDOCK rescoring. ....	213
<b>Figure S2.</b> Preliminary scoring protocols for AMBERDOCK. ....	214

## Introduction

“Facts which at first seem improbable will, even in scant explanation, drop the cloak which has hidden them and stand forth in naked and simple beauty.”

--Galileo Galilei

These words embody the feeling and passion I have for science and discovery. They were written in the introduction of my high school physics textbook, and they have stuck with me ever since. I have always been excited about learning and especially science. Galileo has been referred to as the “father of modern science.” He is the first to clearly state that the laws of Nature are mathematical. In his book *The Assayer* he writes that “the universe...is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures.” The mathematical language of science, for better or worse, helps to explain the most complex inner workings of the Universe. Laws of Nature such as Newton’s laws of motion or Boltzmann’s laws of thermodynamics are immutable and described by simple mathematical equations. Equations such as  $F = ma$  and  $\Delta U = Q - W$ . The very laws that explain how an apple falls from a tree can also be used to explain how the Universe was formed and how life exists.

For the past two years, I have had a plant in my cubicle window. When I received it as a gift, it was in a small pot and only a few inches tall. I have taken good care of it. Fortunately, it does not require a lot of attention. It gets a good amount of morning sunlight, and I water it once a week. Now it is several inches tall and in a pot three times as large. When I water my plant, I admire its earthy smell, and I reflect on how the simple ingredients of dirt, water, air, and light combine to create this complex life. While it has

been a pleasant distraction from long days working at the computer or running experiments in the lab, my plant and others like it offer so much more. Plants, through the process of photosynthesis, convert light energy into chemical energy. They take in atmospheric carbon dioxide and water to make carbohydrates, using sunlight as the energy source. This photosynthetic process is the source of Earth's oxygen. Without it, life as we know it would not exist.

Photosynthesis and all organic and inorganic processes obey the same chemical and physical laws. In 1828 Friedrich Wöhler synthesized the organic compound urea by heating the inorganic compound ammonium cyanate; therefore, showing that compounds found in living organisms could be made in the laboratory from inorganic compounds.<sup>1</sup> The laboratories within living organisms are enzymes. During the 1890's Eduard Buchner and Emil Fischer helped to identify enzymes as the catalysts of biological reactions.<sup>2</sup> Fischer also realized the importance of molecular recognition, which is the idea that only a molecule with a suitable structure can serve as a substrate for a given enzyme.<sup>3</sup> He used a simple analogy of a lock and key to describe this process. Centuries earlier the Roman poet Lucretius spoke of those "things whose textures have such a mutual correspondence, that cavities fit solids, the cavities of the first the solids of the second, the cavities of the second the solids of the first, form the closest union".<sup>4</sup> Drugs ultimately must fit these "cavities" and "solids" to interfere with the natural function of macromolecules such as enzymes and proteins.

The healing effects of certain natural products have been realized for centuries without knowing exactly how they interact in the body. This is still true for many of the drugs used today. In the 5th century BC, Hippocrates wrote about a white powdery substance from willow bark that was known to relieve aches and pains. In 1828 Leroux and Piria isolated the compound salicylic acid, which led to the development of the drug Aspirin in 1899.<sup>5</sup> It was not until 1971 that John Vane discovered that Aspirin, acetylsalicylic acid, suppresses the production of prostaglandins.<sup>6</sup> Prostaglandins are

made by the cyclooxygenase (COX) proteins. The X-ray crystal structures of these were solved in 1994,<sup>7</sup> finally providing a clearer picture of how Aspirin inhibits prostaglandin production. This knowledge has led to the development of many other non-steroidal anti-inflammatory drugs (NSAIDs), a large area of drug research today.<sup>5</sup> I am fond of this story because it was the focus of my first research project as an undergraduate at Vanderbilt and my introduction into structure-based drug design.

Fischer discovered molecular recognition without knowing the structures of the enzymes that he was studying. Even with this amazing foresight, it is unlikely Fischer could have foreseen the evolution of chemistry, biology, and physics into the fields of biochemistry and biophysics of today; nor would he have been able to envision the tools that scientists now take for granted. While X-rays were first discovered in 1895 during Fischer's lifetime,<sup>8</sup> the first protein structures determined to atomic resolution were not solved until the 1960's. At this time, Kendrew solved the structure of myoglobin and Perutz solved the structure of hemoglobin.<sup>9,10</sup> The large amount of data and calculations required to solve these structures required computers. By today's standards, these computers were large and very slow. Computer graphics tools did not yet exist to help visualize the protein structures; rather, wood and metal models were constructed. In his *Nature* paper, Kendrew described a three-dimensional model of myoglobin that he built with 2,500 vertical steel rods at a scale of 5 cm = 1 Å and a total size of six cubic feet.<sup>10</sup> In 1971 the Protein Data Bank (PDB) was founded with seven protein structures.<sup>11</sup> Today there are more than 45,000 deposited structures. Any one of these can easily be downloaded from the Internet and viewed with a multitude of software packages in a matter of minutes.

Scientific models such as Fischer's lock-and-key model or Kendrew's model of myoglobin help visualize or explain complex phenomenon. Models have advanced science and our understanding of the world. Experiments with models often suggest results that are contrary to a scientific hypothesis. The unexpected results or false-

positive predictions often open doors for a whole new area of scientific exploration and understanding. In structure based drug design, structural models of macromolecular receptors are used to design new ligands. The first example of structure-based design occurred in 1973 by Beddell and Goodford.<sup>12</sup> They used the structure of hemoglobin to design ligands that modify oxygen's affinity. Cushman et al. reported the first drug designed by structural information in 1977.<sup>13</sup> Since the 1980's computational methods have been used to help visualize structures of proteins as well as help model and predict receptor-ligand binding.<sup>14, 15</sup> Scoring functions based on physical laws have been developed to capture terms such as electrostatics and solvation.<sup>16-19</sup> Though many make approximations in terms or leave out terms altogether. Computers are considerably faster today compared with those of thirty years ago. However, we still cannot adequately represent the thousands of atoms and millions of pairwise interactions in and between macromolecules, ligands, and water.

When I started my graduate career, most computational methods commonly used to predict receptor-ligand binding focused on reproducing known binding geometries or reproducing known experimental binding data, or both.<sup>20-22</sup> The lack of and inconsistencies in experimental data is a great disadvantage to these docking and scoring methods. Because of their many inadequacies, computational methods at all levels from molecular docking, to MM-GBSA and free energy methods make many false-positive predictions. These come in two forms: compounds predicted to bind, which do not, and compounds for which an incorrect binding geometry is predicted. Respectively, these are referred to as "hit-list" decoys and "geometric" decoys. The hypothesis of this thesis is that these decoys can be used to help pinpoint specific strengths and weaknesses in a given docking and scoring method. Decoys, like failures in life, can be especially informative for making improvements. In addition, these decoys come from several model binding sites which range in complexity from very simple cavity sites to real drug targets. In the simple sites with few isolatable terms, failures are more informative. These

model systems lend themselves to experimental testing. This allows for an iterative cycle of computational prediction, experiment, and algorithm development. Ideally these decoys and model systems will be useful for the computer-aided and structure-based drug design fields. This will help gain a clearer understanding of the physical and chemical principles that govern receptor-ligand binding, which is key to understanding all complex biological phenomena.

In the first chapter, I present a database of decoys for several model and drug-like systems. I examine the strengths and weaknesses of several high throughput molecular docking methods that make many approximations to the energetics of binding. In the second chapter, I used the database of decoys and three cavity sites to test slightly more physically realistic scoring and sampling methods at the MM-GBSA level of theory. Finally, in the third chapter David Mobley and I examined the ability of free energy methods to adequately predict binding free energies for several known ligands; distinguish between known ligands and decoys; and make predictions for five putative ligands. In this work we also looked at taking away energetic sophistication from the free energy method to see which components were the most important for calculating receptor-ligand binding.

## Gloss to Chapter 1.

One of the main reasons I applied to Northwestern University for graduate school was for the opportunity to work in Brian Shoichet's lab. After completing an undergraduate degree in physics and math, I became interested in how these fields could be applied to computational drug research. When I discovered that Brian's lab focused mainly in this area, I knew it would be a perfect fit. Fortunately, he offered me a position as a rotation student in the spring quarter of 2002. My rotation project involved docking and testing compound hits for one of the model drug targets in the lab, AmpC  $\beta$ -lactamase. My rotation mentor, Federica Morandi, had already tested over fifty compounds that were predicted to bind by docking. Out of these hits, she did find one compound that inhibited the enzyme with low micromolar affinity.<sup>23</sup> I then searched for new binding scaffolds that were predicted by docking for this target. I tested more than twenty compounds and found four promising hits. Unfortunately, these hits were actually non-specific aggregators. From a learning standpoint, this project was a great introduction to the computational and experimental methods used in the lab. Though, the results did not seem especially novel from a scientific standpoint. I had no idea at the time that these false-positive docking hits would ultimately shape the remainder of my graduate career.

I officially joined Brian's lab in the summer of 2002. Ken Dill, who is a professor at UCSF with a background in computational methods and their uses in protein folding, initially suggested my first project. Ironically, I would not formally meet Ken until our lab moved to UCSF in the spring of 2003. Ken realized the potential wealth of information that molecular dockers at the time tended to disregard. Molecular docking methods computationally screen thousands to millions of small molecules against a drug target of interest. The hope is to find at least one compound that actually binds with reasonable affinity (mid to low micromolar) to the drug target as predicted. Finding a

lead binder is the first hurdle in the long road towards developing a useful drug.

However, as I learned during my rotation project, the majority of the predicted docking hits fail to bind.

Molecular docking methods rely on their scoring functions to predict ligands. These scoring functions are often flawed, due to the approximations used to make them faster and our lack of knowledge regarding the energetics of ligand binding. Ken suggested developing a database of “geometric” decoys (incorrect geometries of true ligands) and “hit-list” decoys (non-binding ligands predicted to score well) from the model proteins that we had experience with in the lab. Most docking labs developed scoring functions by fitting their scoring terms to reproduce correct binding geometries or known binding energies without explicitly considering decoys.<sup>22, 24-26</sup> The pool of protein-ligand structures and binding data available limited these methods. The protein-folding and protein-protein docking fields had already been using decoy databases to develop their own scoring protocols.<sup>27-30</sup> So why not create a database of small molecule decoys to test and ideally make improvements for molecular docking methods?

I started this project, which resulted in the following chapter, by collecting geometric decoys from five well-studied drug targets. Each of these targets had several known protein-ligand crystal structures available from the PDB. In my search for hit-list decoys, I experimentally tested docking hits that were predicted to bind to three model proteins that the lab was working on at the time. These sites range in complexity from the very simple hydrophobic and polar cavity sites in T4 lysozyme and the complex drug site in AmpC. Finally, my rotation project began to provide results of significant scientific value since I had already discovered twenty decoys for AmpC. Unlike most computational drug designers, I was actually trying to find compounds that would *not* bind to the protein targets. As a final component of the project, I extended the database of decoys from our physics-based method, by rescoring with five other docking scoring functions. Two were knowledge-based scoring functions and three were empirical. The



database of decoys, especially those decoys from the simple cavity sites, opened an avenue that allowed us to start asking specific questions and pinpoint specific problems of fast docking tools, as well as more sophisticated approaches described in Chapters 2 and 3.

Several key conclusions arise from this work. The decoy molecules that make up the high failure rate of docking screens are informative, arguably more so than successful predictions from docking. All six scoring functions that we tested were prone to decoys. The ability to distinguish geometric decoys from native structures was not correlated with performance on hit list decoys. Finally, the model systems lend themselves to simple experiments allowing a cycle of algorithmic development followed by prospective testing. This paper was published in 2005 in the *Journal of Medicinal Chemistry*. In a field that was focused on reproducing known experimental binding energies and previously determined protein-ligand binding geometries, this work was one of the first which highlighted the importance of decoys from experimentally tractable model systems for scoring function testing and development. This paper was cited twice in the *Faculty of 1000 Biology* and is rated a “Must Read”. The supplementary materials are included as Appendix A.

## 1. Decoys for Docking

Alan P. Graves,<sup>1</sup> Ruth Brenk,<sup>2</sup> and Brian K. Shoichet<sup>2\*</sup>

<sup>1</sup>Graduate Group in Biophysics, University of California San Francisco, 600 16<sup>th</sup> St., San Francisco, CA 94143-2240

<sup>2</sup>Department of Pharmaceutical Chemistry, University of California San Francisco, 600 16<sup>th</sup> St., San Francisco, CA 94143-2240

\* Corresponding author:

Phone: 415-514-4126, Fax: 415-514-1460, Email: shoichet@cgl.ucsf.edu

## Abstract

Molecular docking is widely used to predict novel lead compounds for drug discovery. Success depends on the quality of the docking scoring function, among other factors. An imperfect scoring function can mislead by predicting incorrect ligand geometries or by selecting non-binding molecules over true ligands. These false-positive hits may be considered “decoys”. Although these decoys are frustrating, they potentially provide important tests for a docking algorithm; the more subtle the decoy, the more rigorous the test. Indeed, decoy databases have been used to improve protein structure prediction algorithms and protein-protein docking algorithms. Here we describe 20 geometric decoys in five enzymes and 166 “hit list” decoys—i.e. molecules predicted to bind by our docking program that were tested and found not to do so—for  $\beta$ -lactamase and two cavity sites in lysozyme. Especially in the cavity sites, which are very simple, these decoys highlight particular weaknesses in our scoring function. We also consider the performance of five other widely used docking scoring functions against our geometric and hit list decoys. Intriguingly, whereas many of these other scoring functions performed better on the geometric decoys, they typically performed worse on the hit-list decoys, often highly ranking molecules that seemed to poorly complement the model sites. Several of these “hits” from the other scoring functions were tested experimentally and found, in fact, to be decoys. Collectively, these decoys provide a tool for the development and improvement of molecular docking scoring functions. Such improvements may, in turn, be rapidly tested experimentally against these and related experimental systems, which are well behaved in assays and for structure determination.

**Abbreviations:**

L99A, Leu99 → Ala mutant of T4 lysozyme

L99A/M102Q, Leu99 → Ala and Met102 → Gln double mutant of T4 lysozyme

DHFR, dihydrofolate reductase

TS, thymidylate synthase

PNP, purine nucleoside phosphorylase

AChE, acetylcholine esterase

RMSD, root mean square deviation

ACD, the Available Chemicals Directory

CD, circular dichroism

PDB, the Protein Data Bank

HTS, high throughput screening

**Keywords:**

Decoys, molecular docking, virtual screening, cavity, drug design

## Introduction

Molecular docking is widely used to predict novel ligands for molecular targets.<sup>31-</sup>  
<sup>44</sup> In such applications, a large database of organic molecules is screened against a binding site, typically on a protein. These database compounds are often readily available either from vendors or from internal collections. The docked molecules are sampled in multiple conformations and orientations within the binding site, and each configuration is scored for complementarity to the receptor. The best scoring protein-ligand complexes are saved and ranked relative to the rest of the small-molecule database. These best ranking compounds or “hits” can be tested experimentally for binding to the target. Ideally, all would bind with reasonable affinity, but typically most compounds tested fail to bind. In work from this lab for example, 56 compounds predicted to inhibit  $\beta$ -lactamase were tested experimentally, with three of these proving to be true inhibitors. Although often structurally similar to these three novel inhibitors, the other 53 compounds were false positives or “decoys”.<sup>23</sup> Similarly, of 365 molecules predicted as high-ranking hits for PTP1B, 238 (65%) were decoys.<sup>45</sup> This range of hit-rates is not uncommon for the field.<sup>21, 46, 47</sup>

Docking screens have had an impact, notwithstanding these high failure rates, because of their focus on easily available compounds. Thus, whereas the false positives are frustrating, they are *tolerable*. The idea we will develop here is that docking decoys are not only tolerable, but they are actually *useful* for testing and improving docking algorithms. With the right controls and in the right context, they highlight particular weaknesses of an algorithm.

In making this argument, we steal a leaf from work on protein-structure prediction and protein-protein docking.<sup>27-30, 48-55</sup> In these fields, as in small molecule docking and virtual screening, the challenge is to distinguish the native structure from reasonable, but incorrect, alternatives. This is difficult because of the fine balance between solvated and

folded (or bound) states and because of the many configurations and conformations accessible to proteins. Databases of decoy structures have been helpful in refining folding scoring functions by explicitly presenting them with some of the more reasonable of those possible alternative structures. Thus, in protein structure prediction, the Park and Levitt decoy sets,<sup>27</sup> the EMBL decoy sets,<sup>56</sup> and the ROSETTA decoy set<sup>49</sup> are widely used to test new scoring methods. Protein complex decoy sets<sup>55, 57</sup> have been used to a similar effect.<sup>53</sup> The same logic underlying these folding and protein-protein decoys should apply to virtual screening, whose first task is to separate likely geometries and likely molecules from their decoy alternatives.

Because molecular docking aims to identify the correct conformations and orientations of known ligands, as well as predict novel ones, we will consider two types of decoys. The simplest are geometric decoys, where docking predicts an incorrect configuration of a ligand in a binding site. “Hit-list” decoys address the second and arguably more complicated problem of distinguishing true binders from non-binders for a target. These “hit-list” decoys rank highly in docking screens and are predicted to bind, but, on experimental testing, are found not to bind at relevant concentrations.

We will consider geometric decoys for five well characterized enzymes: dihydrofolate reductase (DHFR), thymidylate synthase (TS), purine nucleoside phosphorylase (PNP), acetylcholine esterase (AChE), and thrombin—77 complexes are considered overall. For each system, we find several cases where the docked geometry is correct and several where the best-docked geometry is a decoy. We define a geometric decoy to be a configuration that scores better than the native geometry and that deviates more than 3.0 Å RMSD from the crystallographic configuration thus failing to make key interactions with the binding pocket. For hit-list decoys, we investigate molecules tested as ligands for three well studied binding sites. Two are cavities in the core of T4 lysozyme that are small, well defined, and completely sequestered from bulk solvent. The first of these, created by the substitution Leu99 to Ala (L99A) in the core of the protein,<sup>58</sup>

opens a small, uniformly hydrophobic, solvent inaccessible cavity that binds small aryl hydrocarbons, such as benzene, indene, and naphthalene, but few molecules larger. A second substitution in this site, Met102 to Gln (L99A/M102Q), introduces a single polar atom, the O $\epsilon$  of Gln102, into the cavity. This polar cavity binds, in addition to the apolar aryl hydrocarbons recognized by L99A, more polar molecules such as phenol and aniline derivatives, which do not bind to L99A.<sup>59</sup> The great advantage of these cavity sites is that they are so simple that when a decoy is predicted, the reason it is a decoy is fairly obvious. We will consider 46 decoys for L99A and 24 decoys for L99A/M102Q cavities. Each of these decoys, which scored well by the DOCK3.5.54 scoring function, may be compared to the 56 and 78 known ligands, and the 9 and 12 crystal structures, for the apolar and polar cavities, respectively. Our third model system is a real drug target, AmpC  $\beta$ -lactamase. We will consider 84 decoy molecules predicted for  $\beta$ -lactamase, which may be compared to 26 ligands for this enzyme. In addition, the predictions made for L99A, L99A/M102Q, and  $\beta$ -lactamase can easily be tested experimentally, thus adding to the value of these as model systems for testing and comparing docking algorithms.

Of course, it might be argued that our decoys reflect pathologies of the DOCK scoring function and are not generally interesting for the field. We will therefore evaluate these decoys with five other docking scoring functions including ScreenScore,<sup>60</sup> FlexX,<sup>24</sup> PLP,<sup>25</sup> PMF,<sup>22</sup> and SMOG2001.<sup>26</sup> Whereas DOCK is a force field based scoring function, ScreenScore, FlexX, and PLP are empirical scoring functions which are derived from assigning experimentally determined binding free energies into different additive contributions such as the number of hydrogen bonds, ionic interactions, apolar contacts, and entropy penalties for fixing rotatable bonds in docking the ligand onto the receptor.<sup>61</sup> PMF and SMOG2001 are knowledge-based scoring functions, which use statistical analyses of three-dimensional complex structures to derive a sum of potentials of mean force between receptor and ligand atoms.<sup>61</sup> Brooks et al. carried out a study where they

compared force field, empirical, and knowledge based scoring functions using crystallographic and geometric decoy geometries of 189 protein-ligand complexes.<sup>62</sup> While comprehensive, that study did not include comparisons of scoring functions against virtual screening experiments that include ligands and non-binders or hit-list decoys. Our results support the notion that each of the scoring functions that we tested, including our own, are prone to decoys even against the very simple cavity sites. We will argue that these decoys identify specific problems with each docking scoring function.

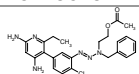
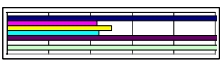
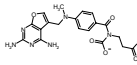
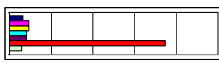
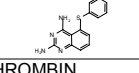
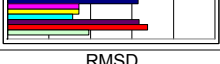
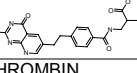
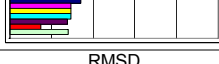
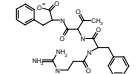

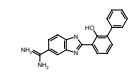
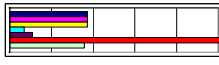
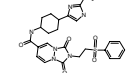
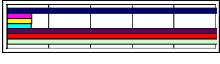
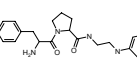

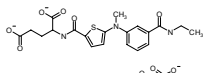

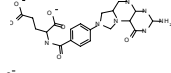
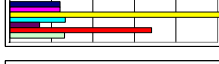
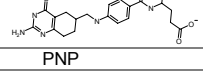
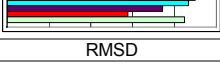
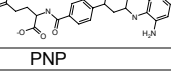
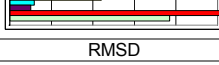
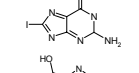

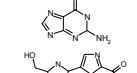

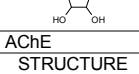
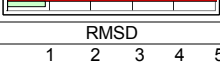
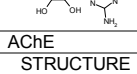
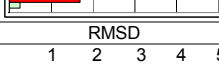
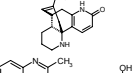
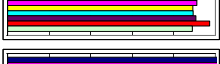
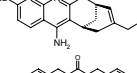

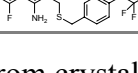

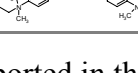
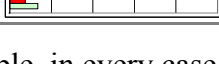


## Results

**Geometric Decoys.** We selected five well-characterized proteins each having several ligand-bound structures in the protein data bank (PDB) to test the ability of a particular scoring function to reproduce the crystallographic or “native” ligand geometries in their cognate proteins. All of the ligands for a particular protein were initially docked against one representative protein structure. These “cross-docking” calculations assume that there is only small conformational change in the protein upon binding different ligands. This rigid treatment of the protein is often used when docking a large compound database. Decoys were also docked to their native protein structures to ensure that they were not simply the product of a “wrong” protein conformation—only decoys that passed this test are listed. When docking against any structure, we also ensured that sufficient sampling of the ligand took place to find poses very close to that determined by crystallography, regardless of their scores (supplementary material). We considered 19 complexes of dihydrofolate reductase, which is a key enzyme in folate biosynthesis; 25 complexes of thrombin, a target for anticoagulant drug therapy; 12 complexes of purine nucleoside phosphorylase (PNP), which is a critical enzyme in the purine salvage pathway; 13 complexes of thymidylate synthase (TS), a well-studied target for anticancer drug design; and eight complexes of acetylcholine esterase (AChE), which is a target for drugs for the management of Alzheimer’s (supplementary material). DOCK3.5.54 was used to generate and score multiple conformations and orientations of each ligand in its cognate protein. In most cases, the best scoring ligand geometries matched the crystallographic ligand geometries to within 2.0 Å root mean square deviation (RMSD); such geometries were considered to be native-like. We focused on ligands that had decoy geometries ( $> 3.0$  Å RMSD from the native pose with better energy scores than any of the native-like dockings) to develop a test set of geometric

decoys. DOCK predicted four geometric decoys for DHFR, five for thrombin, two for PNP, six for TS, and three for AChE. (Table 1 and supplementary material).

**Table 1.** Characteristic geometric decoys and native-like dockings assessed by different scoring functions.

Geometric Decoys						Native-like Dockings									
DHFR			RMSD <sup>a</sup>			DHFR			RMSD						
LIGAND	STRUCTURE		1	2	3	4	5	LIGAND	STRUCTURE		1	2	3	4	5
TABA								MOT							
TQ3								DDF							
THROMBIN			RMSD			THROMBIN			RMSD						
LIGAND	STRUCTURE		1	2	3	4	5	LIGAND	STRUCTURE		1	2	3	4	5
ALG								130							
AIM								PPX							
TS			RMSD			TS			RMSD						
LIGAND	STRUCTURE		1	2	3	4	5	LIGAND	STRUCTURE		1	2	3	4	5
D16								TMF							
DHF								MTX							
PNP			RMSD			PNP			RMSD						
LIGAND	STRUCTURE		1	2	3	4	5	LIGAND	STRUCTURE		1	2	3	4	5
8IG								GUN							
IMR								IMG							
AChE			RMSD			AChE			RMSD						
LIGAND	STRUCTURE		1	2	3	4	5	LIGAND	STRUCTURE		1	2	3	4	5
HUB								HUX							
FBQ								EBW							

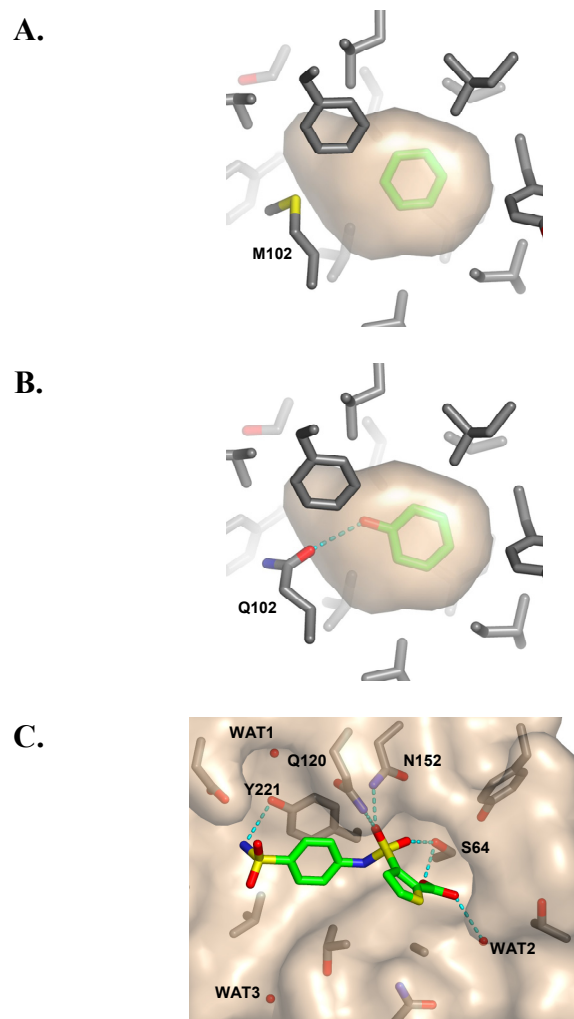
<sup>a</sup> RMSD from crystallographic pose. Although not reported in this table, in every case, the crystallographic pose score for each scoring function was higher (worse) than the energy of the best scoring pose for each scoring function. Blue = DOCK, magenta = ScreenScore, yellow = FlexX, cyan = PLP, purple = PMF, red = SMOG (only those ligands without halogen atoms were rescored by SMOG), and light green = DOCK with an 8-6 van der Waals potential and a higher weighted electrostatics score. A full list of native-like dockings and decoys may be found in the supplementary materials and at <http://shoichetlab.compbio.ucsf.edu/take-away.php>.

To investigate how robust these geometric decoys were, we also evaluated the poses sampled by DOCK3.5.54 with five scoring functions used in molecular docking—ScreenScore, FlexX, PLP, PMF, and SMOG. We used these scoring functions to rescore the predicted geometries for two geometric decoys and two well-matched ligands from each of the five proteins—20 complexes overall (Table 1). Although not reported here, in every case, the crystallographic pose score for each scoring function was higher (worse) than the energy of the best scoring pose for each scoring function; all decoys are scoring decoys, not sampling decoys. In general these scoring functions, with the exception of SMOG, performed no worse than DOCK in those complexes where DOCK found a native-like high-scoring pose. For about half of the geometric decoys found by DOCK, these other scoring functions, again with the exception of SMOG, correctly scored native poses better than decoys (Table 1).

We tested the notion that we could improve DOCK's ability to distinguish native geometries from decoy geometries by softening DOCK's van der Waals potential and by increasing the weight of DOCK's electrostatic score. We softened DOCK's hard 12-6 van der Waals potential to an 8-6 potential to reduce the effect of close contacts between native protein-ligand geometries determined by crystallography. We additionally weighted the electrostatic interaction energy from DOCK by a factor of 4 to simulate the importance of hydrogen bonds. For four out of ten of DOCK's geometric decoys (Table 1), the native geometry was salvaged from the decoy geometries by using the softer van der Waals potential and an increased weight for the electrostatic score. This softer DOCK scoring function only failed on one of the native-like dockings (Table 1). The consequences of this change on hit-list decoys were less promising (below).

**Hit List Decoys.** To investigate hit list decoys, we turned to two well-characterized cavity sites, the L99A and L99A/M102Q lysozyme mutants, and one well-characterized drug target, AmpC  $\beta$ -lactamase (Figure 1). DOCK was used to screen about

200,000 compounds of the Available Chemicals Directory (ACD) against these sites. The screened database contained 49, 70, and 26 known ligands for L99A, L99A/M102Q, and AmpC, respectively.



**Figure 1.** Protein targets used for “hit list” decoys.” (A) Cavity binding site in L99A with benzene (carbons colored green) bound. (B) Cavity binding site in L99A/M102Q with phenol (carbons colored green) bound and forming a hydrogen bond (dashed line) with the O $\epsilon$ 2 oxygen of Gln102. In both A and B the hydrophobic cavity is represented by a tan molecular surface. (C) Active site of AmpC with DOCK predicted pose of ligand 2 (Table 6). The ligand carbon atoms are colored green, three conserved water molecules are represented as red spheres, and hydrogen bonds are drawn with dashed lines. The figures were generated with Pymol (DeLano Scientific LLC, San Carlos, CA).

Of the 49 known ligands for the hydrophobic cavity L99A in the ACD,<sup>59, 63</sup> 39 were predicted by DOCK to score in the top 10,000, which constitutes the top 5%, of the docked database of 200,000 molecules. Their ranks ranged from 21 to 9,880, with 17 in the top 500, or approximately the top 0.25%, of the database (Table 2). There are 45 known non-binders to L99A,<sup>59, 63, 64</sup> 22 of which scored in the top 10,000 with ranks from 46 to 8243 (Table 3). Ten of these scored in the top 500 of the database. There were many others from the top of the hit list that looked either like ligands or non-binders. Of the latter, an additional 8 suspected decoys were tested experimentally and found not to bind detectably to the protein: i.e., were confirmed as decoys (compounds **5-10**, **13**, and **19** in Table 3). Taking into account these new experimental results, a total of 17 decoys scored in the top 500 ranked compounds, and 30 decoys scored in the top 10,000.

**Table 2.** Characteristic L99A Experimentally Tested Ligands Scoring in the Top 10 000 Docking Hits and Their Ranks by Different Scoring Functions.

#	Ligands	Ranking by scoring function <sup>a</sup>						K <sub>d</sub> <sup>b</sup> ( $\mu$ M)	#	Ligands	Ranking by scoring function <sup>a</sup>						K <sub>d</sub> <sup>b</sup> ( $\mu$ M)
		D	S	F	PLP	PMF	S*				D	S	F	PLP	PMF	S*	
1.		<b>21</b>	1156	1112	1961	7847	1826	NA	21.		627	5098	4139	4007	3073	3927	NA
2.		<b>25</b>	949	1120	1273	1312	933	NA	22.		766	932	1494	1303	967	1559	364
3.		<b>33</b>	608	<b>378</b>	1417	2217	1725	102.0	23.		784	1770	3562	1311	1282	<b>236</b>	505
4.		<b>34</b>	<b>341</b>	539	604	1046	793	470	24.		1230	1475	1452	2882	2568	NA	NA
5.		<b>71</b>	885	712	1580	3298	4586	NA	25.		1277	1261	1481	2568	3362	2813	NA
6.		<b>83</b>	<b>323</b>	585	<b>177</b>	<b>344</b>	<b>274</b>	193	26.		2423	3104	4362	3135	2959	<b>356</b>	198
7.		<b>90</b>	1368	616	3739	2682	NA	NA	27.		2432	696	1212	614	3120	1866	NA
8.		<b>152</b>	1625	765	4104	2861	3245	175	28.		2833	4261	5391	1934	1743	1755	NA
9.		<b>182</b>	676	1121	689	906	<b>301</b>	NA	29.		2925	<b>139</b>	<b>193</b>	1153	539	<b>50</b>	NA
10.		<b>218</b>	1304	709	3473	3887	NA	NA	30.		3194	4928	5404	6544	1452	515	120
11.		<b>236</b>	<b>333</b>	<b>498</b>	<b>294</b>	1292	2646	74	31.		3214	1593	1550	3816	2035	NA	NA
12.		<b>302</b>	1076	1346	1776	1345	662	68	32.		3725	2899	3271	4193	1502	<b>156</b>	18
13.		<b>353</b>	5310	4078	4475	6245	6338	NA	33.		3774	6394	7723	3897	2221	1065	NA
14.		<b>363</b>	<b>339</b>	<b>473</b>	<b>310</b>	515	1799	112	34.		4078	3246	2474	5358	1735	NA	NA
15.		<b>365</b>	<b>360</b>	509	<b>325</b>	<b>264</b>	842	290	35.		4289	4055	5935	4892	4237	<b>395</b>	NA
16.		<b>389</b>	565	<b>410</b>	1019	1299	742	NA	36.		4704	3077	4838	3613	4939	1560	NA
17.		<b>420</b>	2464	2245	4099	2901	3563	NA	37.		7787	9691	9424	9558	6699	4885	NA
18.		518	1884	2238	2172	1842	942	422	38.		7923	4396	5597	5470	551	<b>75</b>	14
19.		551	1381	1183	2670	5177	3203	NA	39.		9880	4098	5170	5474	1077	<b>92</b>	19
20.		616	535	743	<b>146</b>	2531	1943	NA									

<sup>a</sup> D = DOCK, S = ScreenScore, F = FlexX, and S\* = SMOG (SMoG ranks are based on a ranking which does not include halogenated compounds). Ranks in **BOLD** font indicate ligands which rank in the top 500 for the respective scoring function. <sup>b</sup> Experimentally determined K<sub>d</sub>'s ( $\Delta$ T<sub>m</sub>'s are known for ligands without a determined K<sub>d</sub>).<sup>63, 64</sup> A full list of L99A ligands may be found in the supplementary materials and at <http://shoichetlab.compbio.ucsf.edu/take-away.php>.



































**Table 3.** Characteristic L99A Experimentally Tested Decoys Scoring in the Top 10 000 Docking Hits and Their Ranks by Different Scoring Functions.

#	Decoys	Ranking by scoring function <sup>a</sup>					S*	#	Decoys	Ranking by scoring function <sup>a</sup>					S*
		D	S	F	PLP	PMF				D	S	F	PLP	PMF	
1.		<b>46</b>	748	685	658	999	NA	16.		<b>371</b>	2235	1418	4613	4802	NA
2.		<b>88</b>	998	441	2554	3374	NA	17.		<b>436</b>	6769	5445	7392	8032	5044
3.		<b>91</b>	869	400	2503	3976	NA	18.		523	2217	3205	<b>229</b>	1883	2608
4.		<b>112</b>	1994	877	4854	4460	NA	19.		607	1137	1872	1139	1950	1172
5.		<b>115</b>	795	557	1393	1473	NA	20.		611	1263	1324	1931	5213	1989
6.		<b>123</b>	1046	889	1788	6035	NA	21.		642	1396	1012	2014	3814	1726
7.		<b>125</b>	8956	6911	9557	9756	NA	22.		671	1220	<b>493</b>	3593	1961	4847
8.		<b>126</b>	1380	1136	2356	4221	3124	23.		807	1635	1250	3594	4138	2077
9.		<b>164</b>	1379	464	4655	1620	NA	24.		1379	4715	5126	2904	4493	3353
10.		<b>175</b>	2610	2271	4031	813	2696	25.		2078	2001	4112	<b>40</b>	<b>107</b>	<b>130</b>
11.		<b>222</b>	737	283	1756	753	NA	26.		2574	2755	3769	3476	580	1431
12.		<b>235</b>	<b>299</b>	<b>110</b>	3127	3801	2641	27.		2694	<b>50</b>	<b>119</b>	<b>264</b>	<b>27</b>	<b>402</b>
13.		<b>249</b>	4761	4086	3214	4545	4403	28.		5504	1804	1152	4670	4172	NA
14.		<b>324</b>	764	<b>349</b>	2387	1952	2559	29.		5936	2319	2774	4565	1388	<b>87</b>
15.		<b>358</b>	832	644	1543	3607	3072	30.		8243	6590	3362	9080	3544	NA

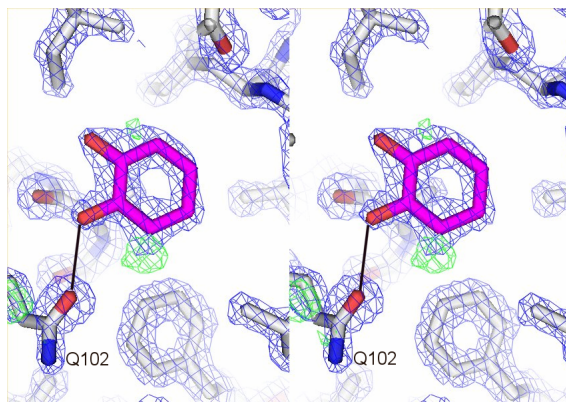
<sup>a</sup>D = DOCK, S = ScreenScore, F = FlexX, and S\* = SMOG (SMoG ranks are based on a ranking which does not include halogenated compounds). Ranks in **BOLD** font indicate decoys which rank in the top 500 for the respective scoring function. A full list of L99A decoys may be found in the supplementary materials and at <http://shoichetlab.compbio.ucsf.edu/take-away.php>.

A slightly more complex cavity is that of L99A/M102Q, which introduces a single polar atom into the otherwise apolar cavity. There were 78 ligands for L99A/M102Q, 55 of which scored in the top 10,000 of the database—in accordance with the observation that the L99A ligands toluene and benzene also bind to the L99A/M102Q site, we assumed that the 56 known ligands of L99A also bind to this more polar site (Table 2 and Table 4).<sup>59</sup> Of these ligands, 15 scored in the top 500, or the top 0.25% of the database (Table 4). There were four known non-binders that scored in the top 10,000,<sup>65</sup> none of which scored in the top 500. Nevertheless, many of the molecules that ranked in the top 500 *looked* like decoys. Seven of these were experimentally tested, and six showed no evidence of binding to the polar cavity (decoys **1-6**, Table 4). Somewhat to our surprise, one compound, catechol, which we thought would not bind owing to excess polarity, does bind to the polar cavity. To understand its basis for binding, we determined the structure of catechol in complex with L99A/M102Q to 1.55 Å resolution by X-ray crystallography (Figure 2). The data suggests two binding modes for catechol. In the first mode, one phenol oxygen of catechol is 2.63 Å and the second is 5.35 Å from the O $\epsilon$  of Q102 as shown in Figure 2. Positive  $F_o-F_c$  density contoured at  $3\sigma$  (green mesh; Figure 2) at the three-position carbon of catechol (Figure 2) suggests a second binding mode in which catechol has rotated 60° counter-clockwise with respect to the first binding mode, and the two phenol oxygens are 2.51 Å and 2.66 Å from the O $\epsilon$  of Q102 (not shown).

**Table 4.** Characteristic L99A/M102Q Experimentally Tested Ligands and Decoys Scoring in the Top 10 000 Docking Hits and Their Ranks by Different Scoring Functions.

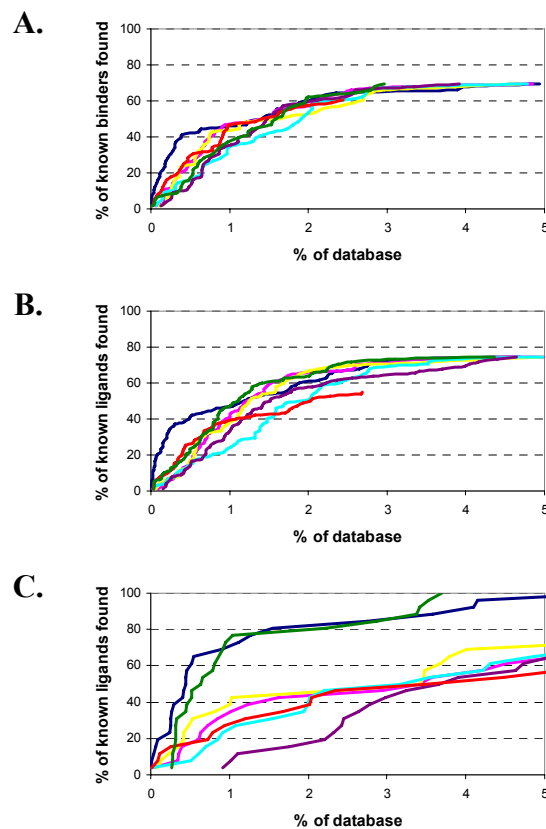
#	Ligands	Ranking by scoring function <sup>a</sup>						K <sub>d</sub> <sup>b</sup> (μM)	#	Ligands	Ranking by scoring function <sup>a</sup>						K <sub>d</sub> <sup>b</sup> (μM)
		D	S	F	PLP	PMF	S*				D	S	F	PLP	PMF	S*	
1.		<b>8</b>	577	<b>332</b>	1200	1027	NA	NA	12.		845	1185	1087	4285	4465	1214	NA
2.		<b>9</b>	361	<b>139</b>	939	5642	NA	NA	13.		979	1419	2765	1081	2017	751	NA
3.		<b>17</b>	1244	1082	2041	2257	2175	156	14.		1052	1067	1373	2956	7399	NA	NA
4.		<b>48</b>	1336	701	3191	8796	NA	NA	15.		1577	1495	1804	3132	2314	NA	NA
5.		<b>60</b>	<b>358</b>	<b>166</b>	780	848	NA	100	16.		2462	1503	1883	2197	1888	NA	NA
6.		<b>171</b>	3278	3708	2640	1430	511	NA	17.		2777	1380	734	5511	9286	NA	NA
7.		<b>308</b>	4738	3169	7045	6720	5371	159	18.		3557	2339	3144	2654	2550	1485	NA
8.		<b>355</b>	1100	708	1857	3676	3857	90.9	19.		4277	1795	2541	4044	1403	<b>81</b>	NA
9.		<b>417</b>	821	942	<b>421</b>	4318	NA	NA	20.		4471	3808	3462	5322	4098	2649	NA
10.		536	3026	2085	5969	5280	NA	56	21.		4593	1150	1147	3436	<b>374</b>	<b>111</b>	NA
11.		606	561	<b>419</b>	1573	4731	3594	NA <sup>c</sup>	22.		5512	2034	1971	5544	2591	<b>323</b>	NA
#	Decoys	D	S	F	PLP	PMF	S*	#	Decoys	D	S	F	PLP	PMF	S*		
1.		<b>28</b>	8531	6677	9212	9803	NA	6.		<b>209</b>	<b>65</b>	<b>43</b>	503	1449	2216		
2.		<b>64</b>	<b>53</b>	<b>18</b>	<b>196</b>	3553	NA	7.		1030	2248	3301	2456	4276	841		
3.		<b>137</b>	4714	4188	2809	3791	5074	8.		1451	1600	2167	1585	1394	744		
4.		<b>152</b>	1740	1876	2304	1323	2970	9.		3261	2908	3447	3057	2291	2446		
5.		<b>198</b>	1581	939	3271	1849	NA	10.		7018	3641	3743	5584	3668	1091		

<sup>a</sup>D = DOCK, S = ScreenScore, F = FlexX, and S\* = SMOG (SMoG ranks are based on a ranking which does not include halogenated compounds). Ranks in BOLD font indicate decoys which rank in the top 500 for the respective scoring function. <sup>b</sup>Experimentally determined K<sub>d</sub>'s.<sup>59</sup> <sup>c</sup>ΔT<sub>m</sub> = 2.6°C. A full list of L99A/M102Q ligands and decoys may be found in the supplementary materials and at <http://shoichetlab.compbio.ucsf.edu/take-away.php>.



**Figure 2.** Catechol bound to L99A/M102Q at 1.55Å resolution. The  $2F_o - F_c$  map is shown in blue wire frame at  $2\sigma$  and the  $F_o - F_c$  electron density map (green) is contoured at  $3\sigma$ . The image was generated with PyMOL (DeLano Scientific LLC, San Carlos, CA).

We were interested in how the other five scoring functions, ScreenScore, FlexX, PLP, PMF, and SMOG, would rank the L99A and L99A/M102Q ligands and decoys. Using these functions, we rescored the top 10,000 ranking compounds against each of the two cavities (Tables 2 through 4). The ranks for 28 out of 39 of the known ligands for L99A that score in the top 10,000 are worsened by three or more of the other scoring functions, as were the ranks of 25 out of 30 of the known decoys. Similarly, when ScreenScore, FlexX, PLP, PMF, and SMOG were used to rescore the top 10,000 scoring compounds against the polar cavity (L99A/M102Q), the ranks of 15 out of 22 of the known binders were lowered by three or more of the scoring functions, as were the ranks of seven out of ten of the decoys. Although the ranks of both ligands *and* decoys were lowered by these other scoring functions, the ranks of the ligands fell further (were ranked worse) than those of the decoys. This is reflected in the overall enrichment factors of the ligands for the different scoring functions against the two cavity sites (Figure 3).



**Figure 3.** Enrichment of ligands for (A) L99A, (B) L99A/M102Q, and (C) AmpC. The percentage of binders found ( $y$ -axis) at each percentage level of the ranked database using the entire ACD ( $x$ -axis). DOCK results are represented by the dark blue line, ScreenScore by magenta, FlexX by yellow, PLP by cyan, PMF by purple, SMOG by red, and the altered DOCK score with a softer 8-6 van der Waals potential and 4-fold increase in electrostatic score is plotted with the green line.

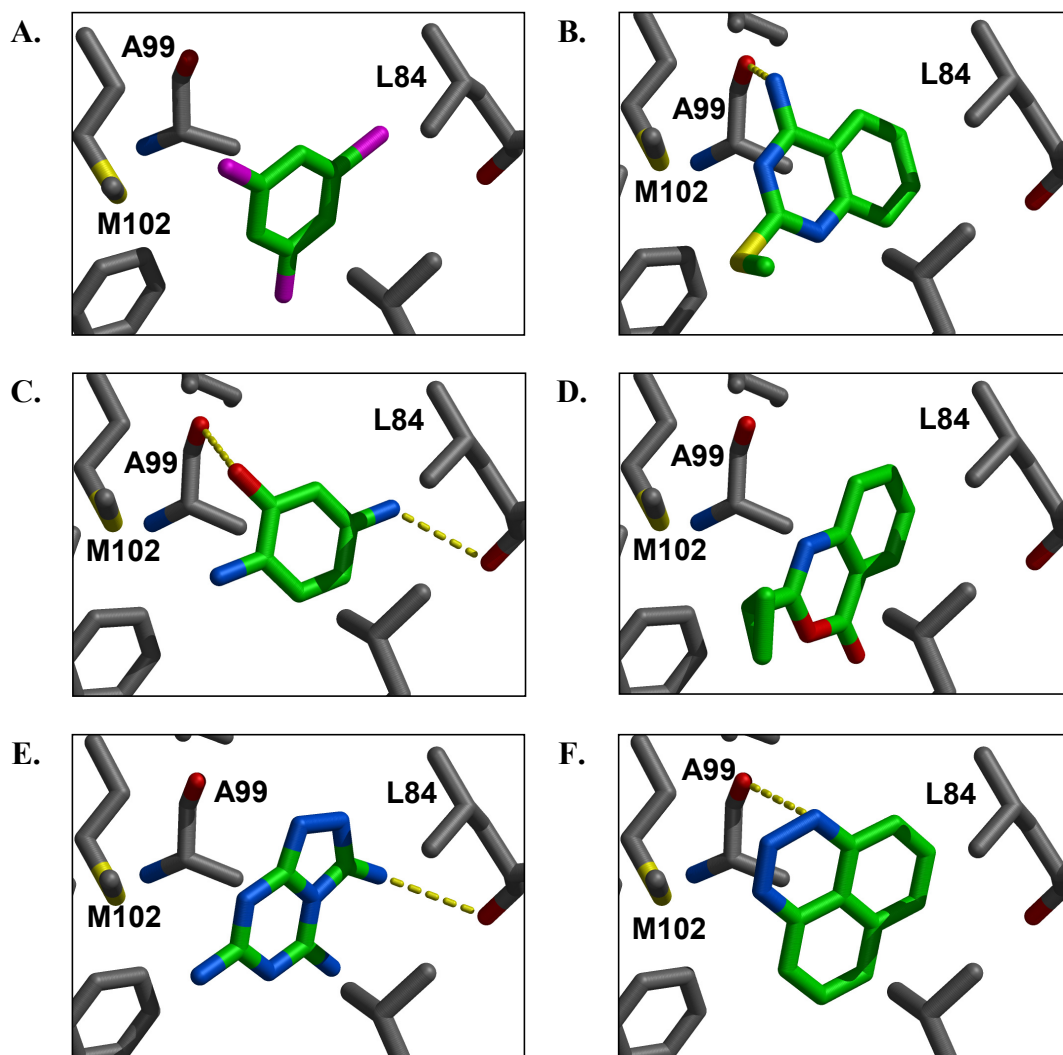
If both ligands and decoys ranked worse by ScreenScore, FlexX, PLP, PMF, and SMOG, a reasonable question is what molecules ranked better? We examined the compounds that ScreenScore, FlexX, PLP, PMF, and SMOG ranked highly (Table 5). To our eyes, the top scoring compounds for these scoring functions typically looked too polar or too large, or both. For instance, many of the very top-scoring molecules for the hydrophobic L99A cavity sported multiple hydrogen bonding groups (Figure 4). Of course, our biases here might be wrong. We therefore tested compounds ranked among the top ten hits for each of the five scoring functions against L99A and L99A/M102Q (17 compounds in total—several were predicted by multiple scoring functions) (Table 5). Of these 17, none were found to bind when tested.

**Table 5.** Decoys for L99A and L99A/M102Q predicted by the ScreenScore, FlexX, PLP, PMF and SMOG scoring functions. All compounds were tested experimentally.

#	Decoys	L99A						#	Decoys	L99A/M102Q					
		Ranking by scoring function <sup>a</sup>								Ranking by scoring function <sup>a</sup>					
		D	S	F	PLP	PMF	S*			D	S	F	PLP	PMF	S*
1.		9419	<b>1</b>	<b>66</b>	<b>2</b>	<b>465</b>	<b>266</b>	1.		571	<b>5</b>	<b>2</b>	<b>288</b>	621	1528
2.		8241	<b>4</b>	<b>12</b>	<b>35</b>	<b>40</b>	2979	2.		987	<b>3</b>	<b>10</b>	<b>112</b>	<b>28</b>	1831
3.		4568	<b>40</b>	<b>150</b>	<b>137</b>	<b>3</b>	<b>435</b>	3.		1941	<b>7</b>	<b>12</b>	<b>41</b>	<b>64</b>	827
4.		7938	<b>9</b>	<b>139</b>	<b>224</b>	<b>21</b>	<b>4</b>	4.		2147	695	1926	<b>1</b>	5606	1718
5.		699	<b>6</b>	<b>1</b>	799	734	1592	5.		2250	<b>24</b>	<b>54</b>	<b>5</b>	<b>57</b>	2169
6.		1972	574	1756	<b>5</b>	6814	2552	6.		2275	<b>1</b>	<b>6</b>	<b>33</b>	<b>10</b>	1933
7.		1424	<b>11</b>	<b>19</b>	<b>56</b>	<b>209</b>	1516	7.		9525	732	596	4709	<b>232</b>	<b>2</b>
8.		7031	<b>111</b>	1349	<b>4</b>	1051	640								
9.		2513	<b>10</b>	<b>7</b>	<b>134</b>	<b>199</b>	2737								
10.		9520	<b>423</b>	694	1993	913	<b>2</b>								
11.		5537	<b>76</b>	<b>280</b>	<b>391</b>	550	<b>3</b>								
12.		9654	1162	3695	<b>328</b>	529	<b>5</b>								

<sup>a</sup>D = DOCK, S = ScreenScore, F = FlexX, S\* = SMOG (SMoG ranks are based on a ranking which does not include halogenated compounds). Ranks in bold font indicate decoys which rank in the top 500 for the respective scoring function.





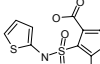
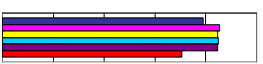

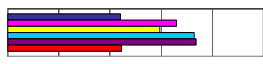
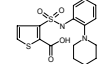
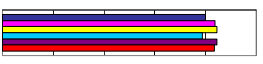


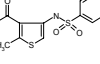



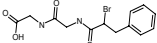

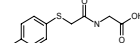

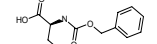

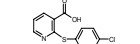

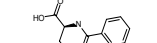

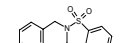

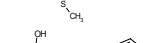

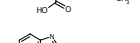

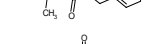
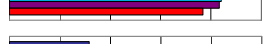
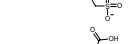

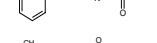

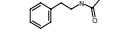

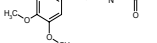

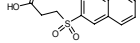

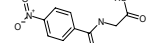

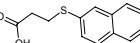

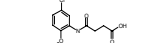



**Figure 4.** Characteristic high scoring docking hits to L99A by (A) DOCK (2<sup>nd</sup> ranking hit), (B) ScreenScore (1<sup>st</sup> ranking hit), (C) FlexX (1<sup>st</sup> ranking hit), (D) PLP (1<sup>st</sup> ranking hit), (E) PMF (1<sup>st</sup> ranking hit), and (F) SMOG (3<sup>rd</sup> ranking hit). The protein carbons are colored gray and the carbons of the docked compounds are colored green. Hydrogen bonds are drawn with dashed lines. The images were generated with MidasPlus (UCSF, San Francisco, CA).

To test the hypothesis that a permissive treatment of steric contacts and an increased emphasis on polar interactions result in worse enrichment of ligands when docking against a large database of decoys, we rescored the top 10,000 hits from both cavity sites by using the altered DOCK score which combined a softened 8-6 van der Waals potential and an increased weight for the electrostatic interaction energy. This scoring function, which had improved performance versus the geometric decoys, enriched fewer ligands for both cavity sites in the top 1% or top 1000 compounds of the database (Figure 3). Compared to the standard DOCK scoring function, the “softened” DOCK scoring function ranked 13 out of 42 L99A and six out of 17 M102Q decoys higher. Beyond the top 1% to 2% of the database, the altered DOCK scoring function improved enrichment of ligands compared to the standard DOCK score. However, this is mostly because we dock against a single, relatively small conformation of the cavities, which cannot easily accommodate some of the larger known ligands in the database without conformational change.

To investigate decoys for a real druggable binding site, we turned to the enzyme  $\beta$ -lactamase, a well-studied target for antibiotic resistance. Unlike the lysozyme cavities, but like most drug targets, the active site of this enzyme presents a mixture of polar and non-polar functionality, is large, and has an extensive solvent interface. There are 26 known non-covalent ligands for AmpC,<sup>23, 66</sup> and 76 known decoy hits from a screen of the ACD database—23 of these decoys were tested for this paper, 53 had been previously discovered (Table 6).<sup>23</sup> All of the ligands and 65 of the decoys scored in the top 20,000, or approximately 10%, from an ACD screen against the AmpC structure. The ligands ranked from 3 to 11,740, with 5 in the top 500 (Table 6). The decoys ranked from 10 to 9344 with 26 in the top 500 (Table 6). Of the 20 high scoring docking hits for AmpC tested for this paper, only one inhibited the enzyme with a  $K_i$  value of about 93  $\mu$ M (ligand 2, Table 6).

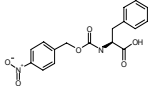

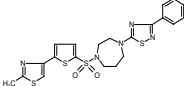

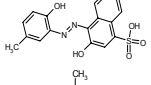

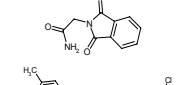

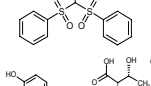

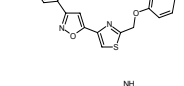

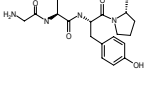
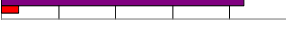
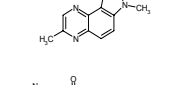

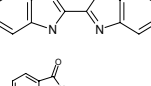

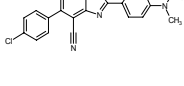

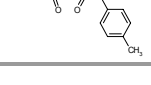

**Table 6.** Charateristic AmpC ligands and decoys and their ranks by different scoring functions.

#	Ligands	RANKS <sup>a</sup>				K <sub>i</sub> <sup>b</sup> ( $\mu$ M)	#	Ligands	RANKS <sup>a</sup>				K <sub>i</sub> <sup>b</sup> ( $\mu$ M)
		10 <sup>1</sup>	10 <sup>2</sup>	10 <sup>3</sup>	10 <sup>4</sup>				10 <sup>1</sup>	10 <sup>2</sup>	10 <sup>3</sup>	10 <sup>4</sup>	
1.						26	5.						64
2.						93 <sup>c</sup>	6.						14
3.						646	7.						60
4.						318							
#	Decoys	RANKS <sup>a</sup>					#	Decoys	RANKS <sup>a</sup>				
		10 <sup>1</sup>	10 <sup>2</sup>	10 <sup>3</sup>	10 <sup>4</sup>				10 <sup>1</sup>	10 <sup>2</sup>	10 <sup>3</sup>	10 <sup>4</sup>	
1.							10.						
2.							11.						
3.							12.						
4.							13.						
5.							14.						
6.							15.						
7.							16.						
8.							17.						
9.													

<sup>a</sup> Rank by scoring function. Blue = DOCK, magenta = ScreenScore, yellow = FlexX, cyan = PLP, purple = PMF, red = SMOG (SMoG ranks are based on a ranking which does not include halogenated compounds). <sup>b</sup> The kinetic data for ligands **1** and **3-7** was reported previously.<sup>23</sup> <sup>c</sup> An apparent K<sub>i</sub> is reported for ligand **2** assuming competitive inhibition and an IC<sub>50</sub> = 240  $\mu$ M. A full list of AmpC ligands and decoys may be found in the supplementary materials and at <http://shoichetlab.compbio.ucsf.edu/take-away.php>.

We used ScreenScore, FlexX, PLP, PMF, and SMOG to rescore the top 20,000 ranking compounds against AmpC (Table 6). As in the cavity sites, the ranks for most of the ligands and decoys were lowered. The ranks for 22 out of 26 of the ligands for AmpC that score in the top 20,000 are worsened by three or more of the other scoring functions, as were the ranks of 45 out of 67 of the decoys (Table 6). We then considered the compounds that ScreenScore, FlexX, PLP, PMF, and SMOG ranked highly (Table 7). As in the cavity sites, we tested compounds ranked among the top ranking hits for each of the five scoring functions. 11 compounds in total—several were predicted by multiple scoring functions—were experimentally tested (Table 7). None of these 11 compounds were found to bind when tested. The “softened” DOCK scoring function was also used to rescore the top 20,000 ranking compounds against AmpC. As in the cavity sites, the less permissive DOCK scoring function enriched fewer known AmpC ligands in the top 1% of the database screen. Figure 3C compares the overall enrichment factors of the known ligands for each scoring function.

**Table 7.** Decoys for AmpC Predicted by the ScreenScore, FlexX, PLP, PMF, and SMOG Scoring Functions.

#	Decoys	RANKS <sup>a</sup>				#	Decoys	RANKS <sup>a</sup>			
		10 <sup>1</sup>	10 <sup>2</sup>	10 <sup>3</sup>	10 <sup>4</sup>			10 <sup>1</sup>	10 <sup>2</sup>	10 <sup>3</sup>	10 <sup>4</sup>
1.			7.								
2.			8.								
3.			9.								
4.			10.								
5.			11.								
6.											

<sup>a</sup> Rank by scoring function. Blue = DOCK, magenta = ScreenScore, yellow = FlexX, cyan = PLP, purple = PMF, and red = SMOG (SMoG ranks are based on a ranking which does not include halogenated compounds).

Two caveats of our results should be considered. First, we only use DOCK generated poses of compounds rather than fully redocking with the other docking programs, which would allow them to generate as well as score ligand poses. It may be that the geometric decoys for these other scoring functions would not have been found if we had allowed them to both sample and score docking poses, because they would have found native-like geometries that DOCK missed. However, we did generate many low RMSD poses regardless of score so at least we can say that many native poses were sampled. Moreover, we note that by and large these other scoring functions did *better* than our own with the geometric decoys. For the hit lists, we only use the other scoring functions to rescore the best scoring DOCK pose. Here too, we know that many true ligands are in the hit lists, in near native geometries, so this is not a question of the right molecules not being available to rank well—they are present. Nor is it a question of gross bias on our part on what may or may not be a ligand or a decoy since several of the best ranking compounds for L99A, L99A/M102Q, and AmpC predicted by the other scoring functions were tested experimentally and found not to bind. The second caveat pertains to how good we are at experimentally distinguishing ligands from decoys. For the two cavity sites, a ligand is a molecule that binds at concentrations of a few millimolar or lower—molecules that might in fact bind at higher concentrations cannot be detected often for solubility or spectral density reasons and so are considered decoys. Similarly, for AmpC, we can detect molecules that bind in the 10mM range, molecules that might bind at higher concentrations will be considered decoys. The range of affinities for known purely non-covalent AmpC ligands is between 1  $\mu$ M and 1 mM. The range for the cavity ligands is between 10  $\mu$ M and about 2 mM.

## Discussion

From a practical standpoint, virtual screening may be considered successful if even 10% of predicted ligands bind to the target at relevant concentrations. From a scientific standpoint, such a failure rate is disconcerting, all the more so since we typically cannot attribute it to a single algorithmic failure. We argue that, with the proper controls and in the proper systems, the decoy molecules that make up the high failure rate of docking screens are informative, arguably more so than successful predictions from docking. Three points stand out from this study. First, all six scoring functions that we tested, including our own, were prone to decoys, often obvious ones. Second, ability to distinguish geometric decoys from native structures was not correlated with performance on hit-list decoys. Third, the model systems discussed here lend themselves to simple experiments, allowing a cycle of algorithmic development followed by prospective testing.

A startling aspect of the decoys is how obvious many of them are. This is most clearly seen in the cavity sites. Molecules like phenol (ranked 235 by DOCK out of over 200,000 molecules docked, decoy 12, Table 3), diaminophenol (ranked 1 by the FlexX scoring function, decoy 5, Table 5), and 8-aminoquinoline (ranked 3<sup>rd</sup> by the PMF scoring function, decoy 3, Table 5) are too polar to bind to the buried, hydrophobic L99A cavity (Figure 1). Molecules like acenaphthylene (ranked 4<sup>th</sup> by the SMOG scoring function, decoy 4, Table 5) are too large for the cavities. That these decoys were, nevertheless, among the very top ranking hits from among the docking scoring functions indicates that they are too permissive to steric violations, desolvation penalties, and frequently both. Why are these violations permitted?

One answer is that these functions may have been devised as initial screens, envisioning more sophisticated secondary calculations to weed out the sorts of decoys that we find here. Thus, a scoring function might be intentionally permissive to steric

violations, implicitly allowing for receptor conformational accommodation that could be properly evaluated with a full energy minimization or molecular dynamics treatment. Such calculations are too costly during a database screen, but might be considered for a smaller list of initial hits. The cost of such permissiveness is to allow decoy molecules as high-ranking hits, to the point that they might crowd out true ligands from the small number of hits possible to re-evaluate with the more sophisticated functions.

Consideration of the performance of the scoring functions on the geometric decoys hints, however, at another explanation for the hit list decoys. Most scoring functions did relatively well on the geometric decoys, distinguishing the native from the decoy poses for most of the 20 complexes we investigated. Docking scoring functions have been extensively tested for their ability to reproduce ligand geometries observed in experimental structures.<sup>21, 62</sup> Indeed, many have been parameterized based on the interactions observed in experimental structures.<sup>22, 24-26</sup> This is similar to protein folding functions parameterized on the interactions in the folded structures of proteins. In folding, it was realized that it is important to consider not only observed interactions but also possible decoy interactions—this has led to the construction of sets of decoy folds by which folding functions are now tested.<sup>27, 28, 48, 49</sup> In small molecule docking, decoys have not been considered in parameterization, at least not formally, and this may have led to an overemphasis on certain interactions and an allowance for certain violations. In parameterizing to reproduce experimental geometries, for instance, one will do well to heavily weight polar interactions, such as hydrogen bonds, which impart directional specificity. Similarly, because steric violations are sometimes present in experimental structures, it is sensible to be permissive to steric repulsion. Such emphasis and allowances can cause problems that only become apparent in a virtual screening application. Whereas polar interactions are key to proper positioning of a molecule, their net contribution to binding affinity is often modest. A scoring function that is heavily biased towards polar interactions may overemphasize polar hits from docking screens,



such as diaminophenol, amino-quinoline, and pterin as ligands for the hydrophobic cavity in lysozyme. Similarly, permissiveness to steric violations will favor larger decoys at the expense of smaller ligands in a database screen.

To test the effect of more sterically permissive and more polar scoring function on geometric and hit-list decoys, we increased the permissiveness to steric violations in the DOCK scoring function and increased the weight of the electrostatic score by 4 fold. This change salvaged four of ten of our geometric decoys (Table 1). Conversely, in database docking against the cavity sites, significantly fewer ligands were found in the top 1000 to 2000 ranking “hits” than were found by the standard, less permissive DOCK scoring function (Figure 3A,B). Thus, whereas a scoring function that is sterically permissive, and that emphasizes polar interactions may do well for reproducing crystal structures, the very same function may do *worse* in database screens.

How extendable are these observations to docking screens against “real” binding sites? The decoys and ligands found for AmpC  $\beta$ -lactamase bear out trends in the toy sites, though admittedly this site, like all real sites, is complex enough to defeat single explanations for decoys. As in the cavity sites, each scoring function predicts several decoys (Table 7). DOCK’s predicted decoys are ranked poorly by most of the other scoring functions (Table 6) and the enrichment of known ligands is worse for the other scoring functions, including the less permissive DOCK scoring function (Figure 3C); alternatively, the other scoring functions have decoys that are ranked poorly by DOCK (Table 7). With the exception of decoys 1 and possibly 4, most of these decoys look unlike the known AmpC ligands. Here too, the decoys are obviously different from the ligands.

Whereas many of the decoys for both the cavity sites and for  $\beta$ -lactamase were obvious, some were fairly subtle. For instance, catechol (ligand 11, Table 4) is a ligand for L99A/M102Q, but 2-aminophenol, which replaces a single hydroxyl group with an amino group, is a decoy (decoy 6, Table 4). We were surprised enough by this difference

to determine the structure of the catechol complex by x-ray crystallography. The electron density of this 1.55 Å structure strongly suggests that catechol has two binding modes in the cavity (Figure 2). Either binding mode in principle would be accessible to 2-aminophenol. A likely reason that 2-aminophenol does not bind is that its amino group is a strong hydrogen bond donor compared to catechol's phenolic oxygens which are fairly weak hydrogen bond acceptors; therefore, the cost of desolvation and binding of 2-aminophenol to M102Q is likely greater than that of catechol. Without experimental binding or structural data, slight differences such as those between catechol and 2-aminophenol can easily be overlooked by even a trained biochemist, not to mention a docking scoring function.

We conclude by returning to the obviousness of many of the hit-list decoys. Whereas this might seem to be a depressing result, we draw some comfort from it. Docking screens have, after all, predicted novel ligands for many receptors,<sup>31-41</sup> notwithstanding their propensity to decoys. What we find encouraging is that fairly simple improvements to docking scoring functions might remove these obvious decoys. Of course, it is possible to treat one type of decoy and introduce another, but in experimentally tractable systems this may be easily tested. We thus hope that the decoy molecules and geometries described here will be useful to the field, leading to a cycle of development and testing in these and other model systems.

## Materials and Methods

**Protein and ligand preparation for single ligand docking.** Ligand bound protein complexes for each of the five enzymes—19 complexes for DHFR, 25 complexes for thrombin, 13 complexes for TS, 12 complexes for PNP, and 8 complexes for AChE—were obtained from the PDB (Table 1, supplementary materials). One representative complex from each enzyme was chosen as the template for docking. 3DFR was chosen for DHFR, 1A4W was chosen for thrombin, 2BBQ was chosen for TS, 1B8O was chosen for PNP, and 1E66 was chosen for AChE. The complexes were then superimposed onto their templates by matching C $\alpha$  backbone atoms of well-defined secondary structural elements. This alignment had no influence on scoring of the docked ligands; it merely simplified the comparison of docked and crystallographic geometries. The resulting matched ligands were then copied into separate files for further preparation. Protons were added to the ligands and atomic partial charges were computed using SYBYL (Tripos, St. Louis, MO). The ligands were converted from pdb to mol2 format. Atom types and bond orders were checked for accuracy, and a docking database for each ligand was prepared from the mol2 formatted ligands. Conformations of each ligand were generated using Omega 0.9 (OpenEye Scientific Software, Santa Fe, NM) and stored in a multiconformer database.<sup>67</sup> Partial atomic charges, solvation energies,<sup>59</sup> and van der Waals parameters<sup>17</sup> were calculated as previously described. The protein structures were prepared for docking as described.<sup>68</sup>

**Molecular Docking of Geometric Decoys.** DOCK3.5.54 was used to dock the ligands to the active site of their respective model proteins. This version of DOCK samples configuration of the ligands more or less finely according to “bin” and overlap distance tolerances.<sup>14, 69</sup> Ligand and receptor bins were set to 0.4-1.0 Å and overlap bins were set to 0.0-0.4 Å; the distance tolerance for matching ligand atoms to receptor

matching sites were set to 1.0-1.5 Å. Each ligand configuration was sampled for steric fit; those passing the steric filter were scored for combined electrostatic and van der Waals complementarity. In any given orientation, the high-scoring ligand conformation was minimized with 20 steps of simplex rigid-body minimization.<sup>70</sup> For each ligand-receptor complex, multiple conformations and orientations of the ligands were written-out. Multiple configurations of twenty of these ligands, four from each enzyme target, were rescored using SCORE and SMOG (see below).

**Docking screens versus L99A and L99A/M102Q cavities and AmpC  $\beta$ -lactamase.** The docking calculations for the cavities were performed as previously described<sup>59</sup> using the benzene bound structure of L99A (181L) and the apo structure of L99A/M102Q (1LGU). The docking database was the 2000.1 version of the ACD (MDL, San Leandro, CA). Compounds containing three or more fluorine atoms as well as compounds containing more than 25 heavy atoms were removed from the database leaving 60,879 molecules in the dockable database. The docking screens for AmpC were performed as previously described<sup>23</sup> using an apo AmpC structure (1KE4). The same version of the ACD was used as the docking database without prior filtering for a total of 220,768 compounds. AMSOL<sup>71, 72</sup> was used to calculate partial atomic charges for each ligand.<sup>59</sup> Conformations of each ligand were generated using Omega 0.9 (OpenEye Scientific Software, Santa Fe, NM) and stored in a multi-conformer database.<sup>67</sup> The best scoring conformation of each of the 10,000 top-scoring molecules against L99A and L99A/M102Q as well as the 20,000 top-scoring molecules against AmpC were saved and rescored using SCORE and SMOG.

**Rescoring the hit lists with SCORE.** Standalone versions of ScreenScore, FlexX, PLP, and PMF scoring functions were implemented in the program SCORE (kindly provided by M. Stahl). SCORE allows one to evaluate any given protein-ligand

configuration by each of these scoring functions. The ligand conformations generated and scored by DOCK3.5.54 were converted to SYBYL mol2 format using an atom typing script in CHIMERA.<sup>73</sup> The bond order information was then added by BABEL version 1.6 (University of Arizona). These scripts simply converted the DOCK output into mol2 format. The SCORE script was then run using the protein pdb file, the active-site pdb file, and a ligand multi-mol2 file to calculate the ScreenScore, FlexX, PLP, and PMF score for each ligand conformation.

**Rescore using SMOG.** Similarly, the docked poses were rescored using the SMOG2001 scoring function (generously provided by B. Dominy and E. Shakhnovich).<sup>26</sup> SMOG uses pdb formatted ligand files, and no additional treatment of DOCK output was necessary. SMOG currently does not have the parameters for halogen atoms so those compounds containing F, Cl, Br, and I were not considered in the enrichment calculations for SMOG.

**Binding of compounds to L99A and L99A/M102Q by upshift of thermal denaturation temperature.** L99A and L99A/M102Q were prepared and purified as described.<sup>59</sup> Thermal denaturation experiments were carried out in a Jasco J-715 spectropolarimeter with a Jasco PTC-348WI Peltier-effect temperature control device and in-cell stirring. To screen the compounds for binding in their neutral forms, denaturation experiments were done at appropriate pH values: compounds 3-fluorobenzonitrile (decoy 5, Table 3), 5-bromopyrimidine (decoy 9, Table 3 and decoy 5, Table 4), and 1,2,4-triazolo[1,5-a]pyrimidine (decoy 9, Table 5) obtained from Aldrich, and 1,6-naphthyridine (decoy 7, Table 5) obtained from TCI were assayed in a pH 5.4 buffer containing 100 mM sodium chloride, 8.6 mM sodium acetate and 1.6 mM acetic acid; compounds 4-vinylpyridine (decoy 10, Table 3 and decoy 4, Table 4), 1-vinylimidazole (decoy 13, Table 3 and decoy 3, Table 4), 2-aminophenol (decoy 6, Table 4), pterin

(L99A decoy 2, Table 5), and 8-aminoquinoline (L99A decoy 2, Table 5) obtained from Aldrich, compound 3,4-diaminofluorobenzene (decoy 2, Table 4) obtained from Avocado Research, compounds 4-amino-2-methylthioquinazoline (L99A decoy 1, Table 5), and 2-aminobenzimidazole (M102Q decoy 3, Table 5) obtained from Acros, compounds 2,5-diaminophenol (L99A decoy 5 and M102Q decoy 1, Table 5) and 7-amino-4-[1,2,4]triazolo[1,5-a]pyrimidin-5-one (M102Q decoy 2, Table 5) obtained from Salor, compounds 4-hydrazinothieno[2,3-d]pyrimidine (L99A decoy 6 and M102Q decoy 4, Table 5) and [1,2,4]triazolo[1,5-a]pyrimidin-7-amine (M102Q decoy 6, Table 5) obtained from Bionet, compound 3-methoxymethylindole (L99A decoy 8, Table 5) from TCI, and adenine (M102Q decoy 5, Table 5) from Sequoia were assayed in a pH 6.8 buffer composed of 50 mM potassium phosphate (a mixture of  $\text{KH}_2\text{PO}_4$  and  $\text{KH}_2\text{PO}_4$ ), 200 mM potassium chloride, and 38% (v/v) ethylene glycol; compounds 2-fluorobenzaldehyde (decoy 6, Table 3), methylchlorodifluoroacetate (decoy 7, Table 3 and decoy 1, Table 4), nitrosobenzene (decoy 8, Table 3), 2-methylbenzyl alcohol (decoy 19, Table 3), catechol (ligand 11, Table 4), acenaphthylene (L99a decoy 4, table 5), 1-naphthalenemethanol (L99A decoy 11, Table 5), 1-methylnaphthalene (L99A decoy 11, Table 5), and 2-benzylpyridine (L99A decoy 12, Table 5) obtained from Aldrich, and compound 2-naphthanitrile (M102Q decoy 7, Table 5) from Acros were assayed in a pH 3 buffer containing 25 mM potassium chloride, 2.9 mM phosphoric acid and 17 mM  $\text{KH}_2\text{PO}_4$ , as described elsewhere.<sup>63</sup>

Thermal denaturation of the protein in the presence of the compounds was monitored by CD at between 223 and 234 nm (although the 223nm wavelength is the ideal wavelength for measuring the helical signal of T4 lysozyme, the higher wavelengths, which were less affected by absorbance from some of the compounds, can be used to monitor the edge of the helical signal). For several compounds with high absorbance in the far UV region, thermal denaturation was monitored by fluorescence emission. Fluorescence was stimulated by irradiation at 280 to 290nm and thermal

denaturation was measured by the intensity of the integrated emission for all wavelengths above 300 nm using a cut-on filter. Thermal melts and data fits were performed as described.<sup>59</sup> Denaturation of the apo L99A and apo L99A/M102Q was performed in the same buffer solutions described above. Potential ligands were included at concentrations between 1mM and 10 mM. Each denaturation experiment was performed at least twice.

**Enzyme Kinetics for AmpC.** AmpC from *Escherichia coli* was expressed and purified to homogeneity as described.<sup>23</sup> Thirty-eight compounds were tested for binding affinity to AmpC. Ligand 2 (Table 6) was obtained from Maybridge. Table 6 decoys 1, 2, 4, and 7 were obtained from Aldrich; decoy 3 from Bachem; decoys 5, 6, 8-11, and 17 from Maybridge; decoys 12, 13, and 15 from Salor; decoy 14 from Buttpark; and decoy 16 from Lancaster. Table 7 decoy 1 was obtained from Pfaltz and Bauer; decoy 2 from Aldrich; decoys 3 and 8 from Salor; decoy 4 from Bachem; decoys 5 and 6 from Asinex; decoys 7 and 9 from Maybridge; decoy 10 from Toronto; and decoy 11 from Bionet. In addition, decoy 29 was obtained from Buttpark; decoy 30 from TCI America; decoys 31-33 from Asinex; decoy 34 from Aldrich; and decoy 35 from Timtec (supplementary material). All were used without further purification. Kinetic measurements with AmpC were performed in 50 mM Tris buffer (pH 7.0) using nitrocefin as a substrate.<sup>23</sup> Reactions were initiated by the addition of enzyme and monitored in methacrylate cuvettes. Any compound showing inhibition was also tested in the presence of 0.01% Triton X-100, to control for promiscuous inhibition.<sup>74, 75</sup> Only ligands that are classic, non-aggregation based inhibitors are reported here.

**Crystallography.** Crystals of the mutant L99A/M102Q were grown using the conditions essentially the same as described,<sup>76</sup> and belong to the space group  $P3_221$ . The crystal was soaked for fifteen minutes in crystallization buffer containing 10 mM catechol. After soaking, the crystal was cryoprotected with Paratone-N (Hampton

Research, Aliso Viejo, CA). X-ray data was collected at 110 K with an in house Raxis IV detector. Reflections were indexed, integrated, and scaled using the HKL package.<sup>77</sup> The complex structure was refined using the CNS package.<sup>78</sup> The X-ray crystal structure has been deposited in the PDB as 1XEP.



## **Acknowledgements**

This work is supported by GM59957 from the NIH (to BKS) and Ernst Schering Research Foundation (to RB). We thank Dr. Martin Stahl for use of the SCORE program, and Dr. Eugene Shakhnovich for use of SMOG. We thank Yu Chen, John Irwin, and Sarah Boyce for reading this manuscript, and Andrej Sali and Matt Jacobson for insightful discussions. We thank Ken Dill for suggesting an investigation of decoy molecules.

## Gloss to Chapter 2.

Chapter 1 describes the database of decoys obtained from several model protein systems. These decoys were predicted by molecular docking methods, which make several approximations or leave out terms that are difficult to calculate. These approximations and missing terms lead to low hit-rates and many false-positive predictions. In the following chapter, the database of decoys and model cavity sites are used to examine more energetically sophisticated scoring methods at the MM-GBSA level of theory. MM-GBSA methods, in contrast to docking, incorporate an improved electrostatic and solvation model as well as improved methods for modeling conformational change. In this work MM-GBSA methods were used to rescore docking hits using a “bottom-up” approach. My collaborators and I believed that by rescoring with a slightly higher level of theory we would see improvements over docking. But more importantly, we wanted to figure out what worked well with MM-GBSA, its limitations, and whether we could apply this information to improve docking.

MM-GBSA methods are orders of magnitude slower than docking and not suited for screening very large databases of small molecules. However, they are an ideal method for rescoring a subset of the best docking hits. While docking generally generates decent compound poses, it often fails to correctly score them. To rescore docking hit lists from three cavity sites, we used two MM-GBSA methods: PLOP from the Jacobson group and AMBERDOCK from the Case group. The polar and hydrophobic sites from T4 lysozyme were discussed in the previous chapter. The third and most complex cavity site, a mutant of Cytochrome C Peroxidase, is anionic with ordered water.

Compared to docking, MM-GBSA methods, with binding site minimization, retrospectively improved the separation of known ligands and known decoys from each of the cavity sites. However, the MM-GBSA methods did not perform well if too much of the site was allowed to relax. The MM-GBSA methods predicted many compounds for

the cavity sites that docking ranked poorly. We prospectively tested 33 of these MM-GBSA hits. Of these, 23 were observed to bind, and most were fairly large for the protein conformation into which they were initially docked. Crystal structures for 21 of these MM-GBSA hits were obtained to compare the computationally predicted poses to the experimentally determined binding geometries. MM-GBSA improved the geometries relative to docking for several of the hits but failed to capture large conformational changes.

While the MM-GBSA methods had improvements over docking methods, some of the high-scoring molecules that it predicted did not in fact bind. These new decoys partly reflect problems in ligand parameterization and partly treatment of electrostatics in the binding pockets. Balancing the opportunities to find new ligands by including receptor relaxation against the potential introduction of new false-positive predictions is a major challenge that this work highlights for MM-GBSA methods and flexible receptor models in general. This work will be submitted to the *Journal of Molecular Biology* in September 2007. Although it is the second chapter in my thesis, this work is the last submitted. While rescoring using subsequently higher levels of theory to filter hit lists is not new,<sup>79-81</sup> this work presents the process using simple model proteins. These systems allow us to ask specific questions as well as make predictions that can easily be tested. Modeling induced fit is currently an area of active research in the field of computational chemistry.<sup>65, 82-85</sup> This work further suggests its importance as well as potential pitfalls. The supplementary materials are included as Appendix B.

## 2. Rescoring Docking Hit Lists for Model Cavity Sites: Predictions and Experimental Testing

Alan P. Graves,<sup>a,b,†</sup> Devleena M. Shivakumar,<sup>c,†,‡</sup> Sarah E. Boyce,<sup>a,d</sup> Matthew P.  
Jacobson,<sup>a,\*</sup> David A. Case,<sup>c,\*</sup> and Brian K. Shoichet<sup>a,\*</sup>

<sup>a</sup>Department of Pharmaceutical Chemistry, <sup>b</sup>Graduate Group in Biophysics, and  
<sup>d</sup>Graduate Group in Chemistry and Chemical Biology, University of California San  
Francisco, 1700 4<sup>th</sup> Street, San Francisco, CA 94158-2330

<sup>c</sup>Department of Molecular Biology, The Scripps Research Institute, 10550 N. Torrey  
Pines Road, La Jolla, CA 92037

<sup>†</sup> These authors contributed equally to this work

<sup>‡</sup> Current address: Institute for Molecular Pediatric Sciences, University of Chicago, 929  
E. 57<sup>th</sup> Street, Chicago, IL 60637

\* Email addresses of corresponding authors:

BKS: shoichet@cgl.ucsf.edu, MPJ: matt.jacobson@ucsf.edu, DAC: case@scripps.edu

## Abstract

Molecular docking computationally screens thousands to millions of organic molecules against protein structures, looking for those with complementary fits. Many approximations are made, often resulting in low “hit rates.” A strategy to overcome these approximations is to rescore top-ranked docked molecules using a better but slower method. One such is afforded by Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) techniques. These more physically realistic methods have improved models for solvation and electrostatic interactions and conformational change compared to most docking programs. To investigate MM-GBSA rescoring, we re-ranked docking hit lists in three small, buried sites: a hydrophobic cavity that binds apolar ligands, a slightly polar cavity that binds aryl and hydrogen-bonding ligands, and an anionic cavity that binds cationic ligands. These sites are simple; consequently incorrect predictions can be attributed to particular errors in the method, and many likely ligands may actually be tested. In retrospective calculations, MM-GBSA techniques with binding site minimization better distinguished the known ligands for each cavity from the known decoys, compared to the docking calculation alone. This encouraged us to test rescoring prospectively on molecules that ranked poorly by docking but that ranked well when rescored by MM-GBSA. A total of 33 molecules highly ranked by MM-GBSA for the three cavities were tested experimentally. Of these, 23 were observed to bind—these are docking false negatives rescued by rescoring. The ten remaining molecules are true negatives by docking and false positives by MM-GBSA. X-ray crystal structures were determined for 21 of these 23 molecules. In many cases, the geometry prediction by MM-GBSA improved the initial docking pose and more closely resembled the crystallographic result; yet in several cases, the rescored geometry failed to capture large conformational changes in the protein. Intriguingly, rescoring not only rescued docking false positives, but also introduced several new false positives into the top-ranking molecules. We

consider the origins of the successes and failures in MM-GBSA rescoring in these model cavity sites and the prospects for rescoring in biologically relevant targets.

**Abbreviations:**

L99A, Leu99 → Ala mutant of T4 lysozyme

L99A/M102Q, Leu99 → Ala and Met102 → Gln double mutant of T4 lysozyme

CCP, Trp191 → Gly mutant of Cytochrome C Peroxidase

MM-GBSA, molecular mechanics with generalized Born surface area approximation

PLOP, Protein Local Optimization Program

RMSD, root mean square deviation

ACD, the Available Chemicals Directory

CD, circular dichroism

UV-VIS, ultraviolet visible

PDB, the Protein Data Bank

HTS, high throughput screening

**Keywords:**

Decoys, molecular docking, virtual screening, MM-GBSA, cavity, drug design

## Introduction

Molecular docking computationally screens large databases of small molecules against a macromolecular binding site of defined structure. The technique is often used to find novel ligands for drug discovery. Notwithstanding important successes,<sup>23, 41, 86-90</sup> docking continues to struggle with many methodological deficits. Many approximations are made to screen many molecules in a timely fashion. These include using only one conformation of the protein, neglecting the internal energies of the docking molecules, using simplified models of ligand solvation energies, typically ignoring protein desolvation, and ignoring most entropic terms entirely. These and other short-cuts lead to the high false positive and false negative rates for which docking screens are notorious. Docking methods are unreliable for affinity prediction and, except in domains of highly related compounds, even for rank ordering the likely hits that emerge from the virtual screens.

To overcome these deficits, several groups have combined disparate scoring functions in a consensus fashion to capitalize on the strengths and overcome the deficiencies of individual methods.<sup>91-93</sup> This “consensus scoring” approach is attractive when it has worked, but its theoretical underpinnings are slim.<sup>94</sup> An alternative approach involves using a higher level of theory to re-score the docking hit lists after the docking calculation has completed. The goal is to re-evaluate the top docking hits for energetic complementarity to the target after including more terms and degrees of freedom than modeled by the docking program. Because more terms are considered, rescoring is typically much slower than docking, so much so that only the top-scoring docking pose of the best scoring docked molecules are often considered. This approach has been adopted by versions of the program GLIDE.<sup>79</sup> Here ligands are first docked using simplified and relaxed criteria and are then refined by more sophisticated and stringent evaluation of the energies of binding. Similarly, Kollman used a hierarchical technique



that begins with initial database screening and progresses to Molecular Mechanics-Poisson-Boltzmann Surface Area (MM-PBSA) rescoring to find HIV-1 Reverse Transcriptase inhibitors.<sup>80</sup> The combination of an initial docking screen with subsequent re-scoring by a Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) method has been used to improve enrichment of known ligands for several enzymes in retrospective studies and even to identify substrates.<sup>95-99</sup>

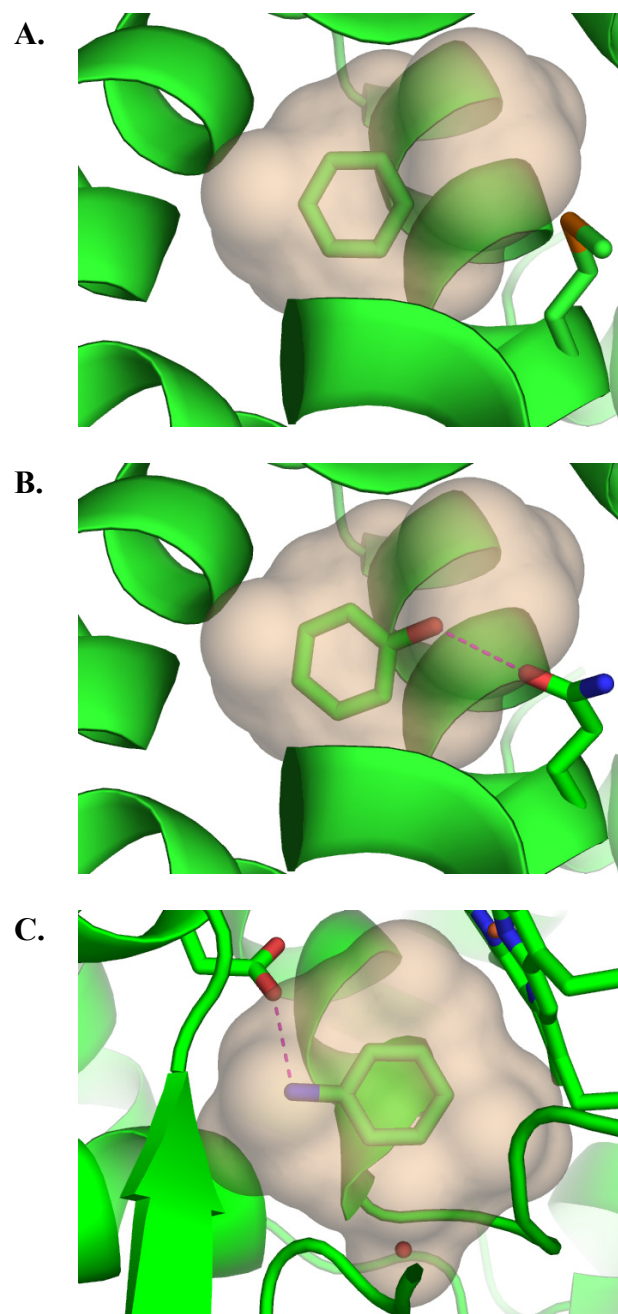
Such MM-PBSA and MM-GBSA methods involve minimization and often dynamic sampling of the protein-ligand complexes, and include ligand and receptor conformational energies and strain. They evaluate the electrostatics and solvation components of the binding energy by PB or GB methods, including both ligand *and* receptor desolvation. The MM-GBSA binding energy is determined by

$E_{complex} - E_{receptor} - E_{ligand}$  where  $E$  is an MM-GBSA estimate and solute configurational entropy effects are ignored. In this paper, we focus on relative binding energies of different ligands to the same receptor, so the free receptor energy ( $E_{receptor}$ ) does not affect the results. Because the MM-GBSA function includes both internal energies and solvation free energies, and because we explicitly subtract complex ( $E_{complex}$ ) and ligand ( $E_{ligand}$ ) contributions, desolvation effects upon complex formation for both the ligand and the receptor are included, at least in principle. There are three main limitations: (1) the force fields and solvation energies are not uniformly accurate; (2) for reasons of computational efficiency, only a small part of configuration space near the DOCK starting pose is really explored; and (3) configurational entropy effects are ignored. Notwithstanding these limitations, the MM-GBSA methods represent a substantially higher level of theory than that encoded by most docking programs and are attractive alternatives to a more complete treatment of the energies of interaction by free energy perturbation (FEP) and thermodynamic integration (TI),<sup>100</sup> which remain the gold standard but are very slow.

In this study, we set out to test MM-GBSA rescoring of docking hit lists in simple model cavity sites. These sites have been engineered into the buried cores of proteins and bind multiple small organic molecules. In contrast to most drug targets, these cavities are small (150-180 Å<sup>3</sup>), buried from bulk solvent, and are dominated by a single interaction term. The L99A (Leu99→Ala) cavity in T4 lysozyme<sup>63</sup> is almost entirely apolar, the L99A/M102Q (Leu99→Ala/Met102→Gln)<sup>59</sup> cavity in the same protein has a single hydrogen-bond acceptor (the introduced Gln102), whereas the W191G (Trp191→Gly) cavity in Cytochrome C Peroxidase (CCP)<sup>101, 102</sup> has a single anionic residue, Asp235 (Figure 1). The ligands recognized by these sites correspond to these features: the hydrophobic L99A binds small, typically aromatic non-polar molecules; the slightly polar L99A/M102Q binds both apolar molecules but also those bearing one or two hydrogen-bond donors; whereas, the anionic W191G cavity almost exclusively binds small monocations. The simplicity of these sites is conducive to disentangling the energetic terms of ligand binding, which are so often convoluted in drug targets with their larger, more complex binding sites. It should be noted that previous work with solvent exposed sites has suggested that a major advantage of MM-GBSA scoring functions is calculating partial receptor desolvation upon ligand binding.<sup>96</sup> This benefit with complex solvent exposed binding sites may be less relevant in the buried cavity sites, especially the hydrophobic L99A and polar L99A/M102Q sites, which are mostly desolvated. (It is our experience that the cavity sites, in fact, impose a greater strain on the GBSA solvent models to fully desolvate the pockets.)

In the cavity sites, an incorrect prediction is often informative, identifying a single problematic term in a scoring function; we have used these cavities as model binding sites to identify problems in molecular docking<sup>59, 65, 103, 104</sup> and, more recently, thermodynamic integration.<sup>100</sup> Others have found them attractive test systems for methods development studies.<sup>105-108</sup> An important advantage of these cavity sites is that they are experimentally tractable for detailed, prospective testing of ligand predictions.

Because the ligands they bind are small—in the 70 to 150 amu range—many possible ligands are readily available commercially, which is rarely true of drug targets.<sup>109</sup> The binding of these predicted ligands may be tested by direct binding assays, and the structures of the ligand-protein complexes may be routinely determined by x-ray crystallography to resolutions better than 2 Å. Extensive study in the Matthews, Goodin, and our own laboratories has resulted in many tens of diverse ligands for each cavity, as well as tens of “decoys,” which are molecules that were predicted to bind to the sites but for which no binding was observed at concentrations as high as 10 mM on experimental testing.<sup>59, 65, 100, 103, 104</sup>



**Figure 1.** The model cavity sites. A. Cavity binding site in T4 lysozyme L99A with benzene bound. B. Cavity binding site in T4 lysozyme L99A/M102Q with phenol bound; the hydrogen bond with the O $\epsilon$ 2 oxygen of Gln102 is represented by a dashed line. C. Cavity binding site of cytochrome C peroxidase W191G with aniline bound; the hydrogen bond with Asp235 is represented by a dashed line. The heme and an ordered water molecule are also depicted. In A., B., and C. the cavities are represented by a tan molecular surface and the protein ribbons are colored green. Rendered with the program PyMOL.<sup>110</sup>

We thus used these three simple model cavity sites, L99A, L99A/M102Q, and W191G, as templates to measure the strengths and weaknesses of MM-GBSA rescoring of docking hit lists. We used two rescoring programs: PLOP<sup>111, 112</sup>, with binding site side chain rotamer search and minimization, and AMBERDOCK, using short MD steps and minimization of binding site residues (Materials and Methods). Molecular docking was used to screen compound libraries that contained between 5000 and 60,231 fragment-like molecules from the Available Chemicals Directory (ACD). The single best pose for each compound that ranked among the top 5000 or 10000 compounds by docking was then rescored by both MM-GBSA programs. Multiple known ligands and decoys were among the molecules rescored for all three sites' rescored sets. In retrospective calculations, MM-GBSA rescoring improved the separation of ligands from decoys in each of the cavities. We then tested 33 new ligands that were predicted to bind by the MM-GBSA methods that docking alone ranked poorly—generally much worse than the top 500. To investigate the detailed basis of the MM-GBSA predictions, we determined crystal structures for 21 of these new ligands and compared them to the geometries predicted by theory. These studies suggest areas where MM-GBSA methods can contribute to the success of virtual screening, and areas where this method faces important challenges.

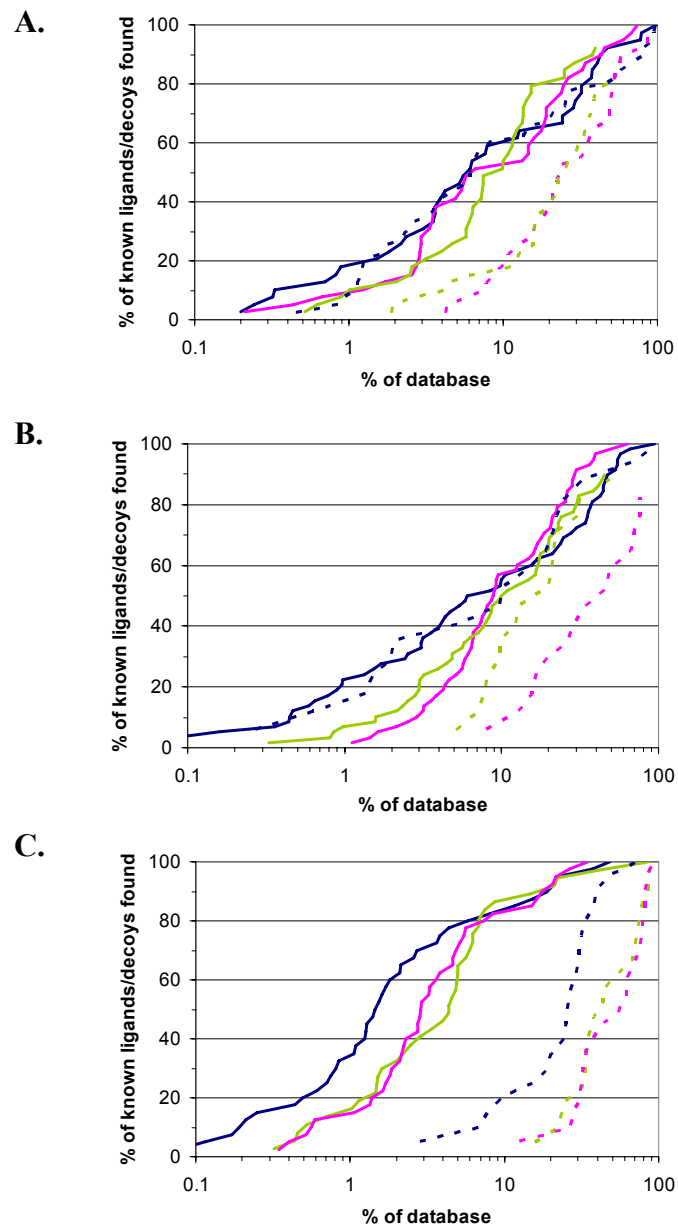
## Results

### Retrospective Docking and Rescoring in the Hydrophobic Cavity.

Approximately 60,000 small molecules were docked into the hydrophobic cavity L99A using DOCK3.5.54<sup>59, 67</sup> (Figure 1a). Among the top-scoring 10,000 molecules were 39 known ligands and 40 experimentally tested decoys. DOCK found 44% (17 molecules) of these ligands and 43% (17 molecules) of these decoys among the top 500 molecules (Figure 2a). Ligands such as toluene (DOCK rank 32), benzene (DOCK rank 151), and ethylbenzene (DOCK rank 301) are small, aromatic and hydrophobic compared to known decoys such as nitrosobenzene (DOCK rank 125), phenol (DOCK rank 234), and 3-methylpyrrole (DOCK rank 435). Like the ligands, these decoys are also small and aromatic, but are presumably too polar for the hydrophobic cavity to overcome their desolvation penalty (Figure 1a).

The top-ranking 10,000 docking hits for the hydrophobic cavity were re-ranked by PLOP and the top-ranking 5,000 docking hits were re-ranked by AMBERDOCK. For both methods, the enrichment of the ligands actually decreased slightly relative to that achieved by docking alone; that is to say, fewer ligands were found among the very best scoring molecules (Figure 2a). Rescored by PLOP, 41% (16 molecules) of the known ligands were found among the top 500 molecules, whereas 28% (11 molecules) were found by AMBERDOCK. Both enrichment factors were lower than those found by docking alone. On the other hand, the enrichment of the known decoys was lower still (Figure 2a). Only 5% of the decoys (2 molecules) were ranked among the top 500 molecules by PLOP and only 13% (5 molecules) were so ranked by AMBERDOCK. This represents a substantial improvement on docking alone. We should note that both the ligand enrichment and the decoy enrichment are strongly biased for docking—many of the ligands and almost all of the decoys were originally tested based on docking predictions<sup>59, 65, 104</sup>—so it is reasonable to expect that the enrichment of ligands will be

higher by docking, as will the decoys. Perhaps more informative than is the separation of the ligands from the decoys, as measured by the ratios of their enrichment factors. These were improved eight-fold by PLOP and two-fold for AMBERDOCK, relative to that of DOCK in this hydrophobic cavity.



**Figure 2.** Retrospective enrichment of ligands and decoys for A. the hydrophobic L99A cavity, B. the polar L99A/M102Q cavity, and C. the anionic W191G cavity. The plots depict the percentage of known ligands (solid lines) or decoys (dashed lines) found (*y*-axis) at each percentage level of the ranked database using the top 10,000 best scoring docking hits (*x*-axis) for L99A (A) and L99A/M102Q (B) and the 5400 best scoring docking hits (*x*-axis) for CCP (C). Docking enrichment of known ligands (solid) and decoys (dashed) are represented by the dark blue curves. PLOP enrichment of known ligands (solid) and decoys (dashed) are represented by the pink curves. AMBERDOCK enrichment of known ligands (solid) and decoys (dashed) are represented by green curves.



**Retrospective Docking and Rescoring in the Polar Cavity.** The same 60,000 molecules were docked into the polar cavity L99A/M102Q (Figure 1b). Among the top-scoring 10,000 molecules were 58 ligands and 17 experimentally tested decoys. DOCK found 45% (26 molecules) of these ligands and 35% (6 molecules) of these decoys among the top 500 molecules (Figure 2b). The increased polarity from O $\epsilon$  of the Gln102 side chain in the cavity accommodates the binding of phenol (DOCK rank 354) and 3-methylpyrrole (DOCK rank 307), which are decoys for the L99A cavity, as well as hydrophobic ligands such as toluene (DOCK rank 16) and benzene (DOCK rank 78). The increased polarity of the site only goes so far, however, and it cannot accommodate decoys such as 1-vinylimidazole (DOCK rank 136) or 2-aminophenol (DOCK rank 208), whose polarity is presumably still too great for the single carbonyl oxygen of the site to overcome the attendant desolvation terms.

The top 10,000 docking hits for the polar cavity were re-ranked by PLOP and the top 5,000 re-ranked by AMBERDOCK. For both methods, the enrichment of the ligands again decreased slightly relative to the docking enrichment factor (Figure 2b). Rescored by PLOP, 22% (13 molecules) of the known ligands were found among the top 500 molecules, whereas 34% (20 molecules) were found by AMBERDOCK. However, the enrichment of the known decoys was lower still. None of the decoys were ranked among the top 500 molecules by PLOP or AMBERDOCK, in contrast to DOCK where 35% (6 molecules) of the known decoys were scored among the top 500 molecules. As in the hydrophobic site, despite the decrease in overall ligand enrichment, the separation of the ligands from the decoys was improved substantially for the polar cavity: by 20 fold for PLOP and 35 fold for AMBERDOCK.

**Retrospective Docking and Rescoring in the Anionic Cavity.** Approximately 5400 molecules were docked in the charged cavity of CCP (Figure 1c). Within this database were 40 known ligands and 20 experimentally tested decoys. DOCK found 78%

(31 molecules) of these ligands and 20% (4 molecules) of these decoys among the top 500 molecules (Figure 2b). The anionic cavity typically binds cationic ligands such as 2-aminopyridine (DOCK rank 6) and imidazole (DOCK rank 227). Most neutral polar compounds, such as 3,5-difluoroaniline (DOCK rank 148), and apolar compounds, such as toluene (DOCK rank 411), are decoys for this cavity, as are anionic compounds or those bearing a formal charge greater than +1.

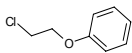
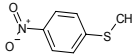
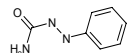
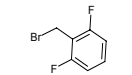
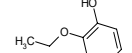
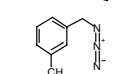
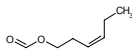
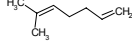
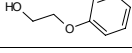
All of the 5400 docking hits for the anionic cavity were re-ranked by PLOP and AMBERDOCK. Rescored by PLOP, 83% (33 molecules) of the known ligands were found among the top 500 molecules, and 80% (32 molecules) were found by AMBERDOCK (Figure 2c). Both enrichment factors are comparable to those found by docking alone, which found 83% (33 molecules) of the known ligands among the top 500 molecules. On the other hand, fewer of the known decoys were enriched by the MM-GBSA methods. None of the known decoys were ranked among the top 500 molecules by PLOP or AMBERDOCK, and the best scoring decoy ranked 655 for PLOP and 785 for AMBERDOCK compared to 145 for docking. Thus, whereas the overall enrichment of the ligands relative to the rest of the database molecules remained unchanged, the separation of the ligands from the decoys was improved by four-fold for PLOP and AMBERDOCK.

**Prediction and Experimental Testing of New Ligands.** A more robust test, one less biased by previous knowledge, involves prospective prediction of new ligands. For each of the three cavities, we looked for molecules that had been poorly ranked by docking but that ranked well by either PLOP or AMBERDOCK or both. Nine compounds were picked and tested for the hydrophobic L99A cavity, ten were tested for the polar L99A/M102Q cavity, and fourteen were tested for the anionic W191G cavity. Structures for 21 of these 33 molecules in complex with the cavities were determined by protein crystallography, allowing us to compare the predicted and experimental

geometries in detail. In the following discussion, we report whether binding was detected at a single concentration tested. The actual affinities were not measured but will often be substantially better than the concentration reported.

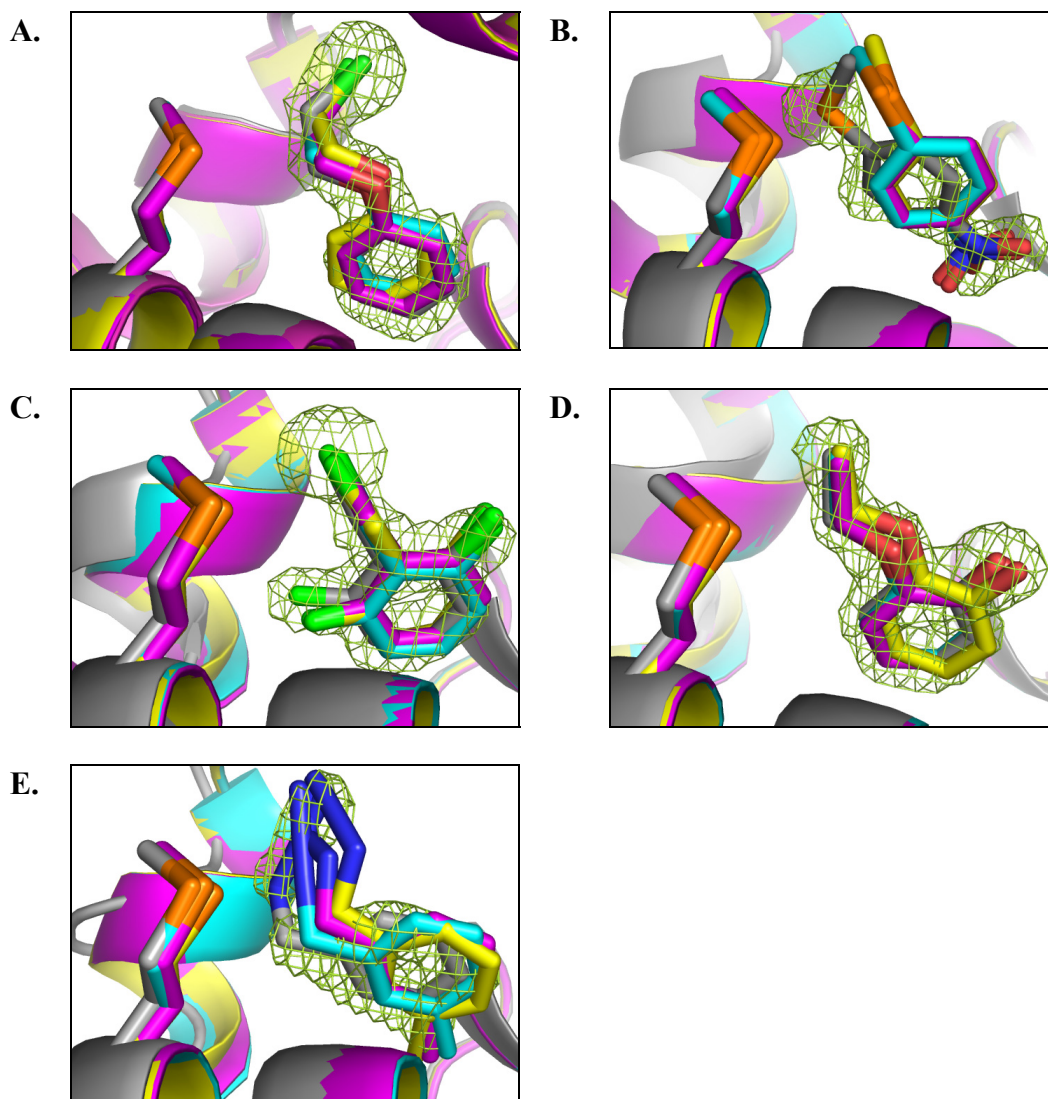
**New L99A Ligands Predicted by Rescoring:** All of the nine ligands predicted by PLOP and AMBERDOCK were relatively large compounds that do not easily fit into the unminimized cavity into which they were docked, explaining their poor docking ranks, but they fit well upon receptor relaxation by MM-GBSA. Binding was detected at millimolar concentrations by temperature of melting ( $T_m$ ) upshift experiments for seven of these nine compounds; however, for two no binding was detected (Table 1). AMBERDOCK correctly predicted binding for five ligands and incorrectly predicted binding for 1-phenylsemicarbazide and 2-phenoxyethanol (two of the prospectively tested molecules were not rescored by AMBERDOCK because docking ranked them worse than 5000). PLOP correctly predicted binding for five ligands, while incorrectly predicted binding for 2-phenoxyethanol. PLOP agreed with docking on the remaining three molecules that had been prioritized by AMBERDOCK, ranking them worse than 1000. Two of these, 4-(methylthio)nitrobenzene and 2-ethoxyphenol, were true ligands and so are false negatives for PLOP.

**Table 1.** Compounds predicted by AMBERDOCK and PLOP to bind to T4 Lysozyme L99A.

Structure	Compound	Score and Rank <sup>a</sup>			C  <sup>b</sup> (mM)	pH	$\Delta\Delta H$ (Kcal/mol)	$\Delta T_m$ (°C)	Binding detected	Structure determined
		DOCK	AMBER	PLOP						
	$\beta$ -Chlorophenetole	-4.89 (3786)	<b>-22.38</b> <b>(5)</b>	<b>-26.31</b> <b>(15)</b>	10	3.0	31.0	6.5	Yes	Yes
	4-(Methylthio)nitrobenzene	-5.69 (3358)	<b>-22.36</b> <b>(6)</b>	-16.22 (1243)	<10	3.0	6.2	1.3 <sup>c</sup>	Yes	Yes
	1-Phenyl-Semicarbazide	-4.49 (3965)	<b>-22.03</b> <b>(8)</b>	-7.69 (5290)	10	6.8	1.6	-0.9 <sup>c</sup>	No	NA
	2,6-Difluorobenzyl Bromide	-10.59 (1046)	<b>-22.01</b> <b>(9)</b>	<b>-21.10</b> <b>(186)</b>	<10	3.0	10.0	1.6 <sup>c</sup>	Yes	Yes
	2-Ethoxyphenol	-6.74 (2806)	<b>-21.54</b> <b>(12)</b>	-15.19 (1642)	5	3.0	12.0	1.2	Yes	Yes
	3-Methyl-Benzylazide	-10.54 (1061)	<b>-19.58</b> <b>(57)</b>	<b>-25.19</b> <b>(27)</b>	10	3.0	8.1	1.5	Yes	Yes
	cis-3-Hexenyl-Formate	3.61 (7746)	NR <sup>d</sup>	<b>-25.17</b> <b>(28)</b>	10	3.0	5.1	1.5	Yes	NA
	6-Methyl-1,5-Heptadiene	1.78 (7035)	NR <sup>d</sup>	<b>-24.92</b> <b>(30)</b>	10	3.0	14.9	2.8	Yes	NA
	2-Phenoxyethanol	-5.76 (3323)	<b>-19.64</b> <b>(55)</b>	<b>-23.13</b> <b>(68)</b>	10	3.0	2.7	0.0	No	NA

<sup>a</sup> Compound scores and ranks (in parenthesis) for DOCK, AMBERDOCK, and PLOP. Scores and ranks in bold font indicate ligands which rank in the top 200 for the respective scoring function. <sup>b</sup> Concentration at which ligand was tested. <sup>c</sup>  $\Delta T_m$  monitored using fluorescence, exciting at  $\lambda=283\text{nm}$  and measuring the integrated emission above 300 nm. <sup>d</sup> NR is not ranked.

Five high resolution (better than 2 Å) protein-ligand crystal structures were obtained for these new L99A ligands to compare experimental to predicted poses (Figure 3). In each case, electron density for the ligands was unambiguous, allowing us to model their positions in the site. Docking and MM-GBSA methods predicted the binding geometry for three of the five ligands to within 0.3 to 0.8 Å RMSD (Table 2). Conversely, the docked pose of 3-methylbenzylazide was 1.4 Å RMSD from the crystallographic pose. The PLOP minimized prediction had a slightly improved RMSD of 1.1 Å, but the refined ligand also had a non-linear azide group, highlighting a failure in ligand parameterization. In addition, docking and MM-GBSA methods predicted poses which were approximately 1.5 Å RMSD from the crystallographic pose of 4-(methylthio)nitrobenzene. The crystallographic poses of these two ligands would have been within 2 Å of the Val111 side chain in the conformation of the cavity used for the docking calculation, a steric conflict that is relieved by conformational expansion of the cavity in the experimental structures. Indeed, for all complexes, with the exception of  $\beta$ -chlorophenetole, the F-helix of lysozyme (residues 108-113) that forms one wall of the cavity reorients by about 2 Å and swings Val111 further out of the cavity to accommodate the ligands.<sup>113</sup> The protein conformations seen in these structures more closely resemble the larger isobutylbenzene bound cavity site (PDB id 184L) than the smaller benzene bound cavity site (PDB id 181L) used for docking and rescoring. Whereas the MM-GBSA methods do not capture this helix motion, receptor and ligand minimization reduces the steric clash sufficiently to improve the ranks of what were docking false negatives. Higher level calculations using free energy methods and molecular dynamics have captured the F-helix motion and explained discrepancies in free energies upon ligand binding due to its displacement.<sup>100, 105</sup>



**Figure 3.** Predicted and experimental ligand orientations for the hydrophobic L99A cavity. The carbons of the crystallographic pose, the DOCK predicted pose, the AMBERDOCK predicted pose, and the PLOP predicted pose are colored grey, yellow, cyan, and magenta, respectively. The  $f_o-f_c$  omit electron density maps (green mesh) are contoured at  $2.5-3.0\sigma$ . A.  $\beta$ -chlorophenetole, B. 4-(methylthio)nitrobenzene, C. 2,6-difluorobenzylbromide, D. 2-ethoxyphenol, and E. 3-methylbenzylazide bound to L99A. Rendered with the program PyMOL.<sup>110</sup>

**Table 2.** Crystallographic measurement and the RMSD values for predicted and crystallographic ligand geometries in the L99A and L99A/M102Q sites.

	L99A Ligands					L99A/M102Q Ligands					
	$\beta$ -chloro phenetole	4-(methylthio) nitrobenzene	2,6-difluoro benzyl bromide	2-ethoxy phenol	3-methyl benzylazide	n-phenyl glycino nitrile	2-nitro thiophene	2-(n-propyl thio)ethanol	3-methyl benzylazide	2-phenoxy ethanol	3-chloro-1-phenyl-1-propanol
Resolution (Å)	1.80 (1.84)	1.64 (1.68)	1.84 (1.89)	1.70 (1.74)	1.46 (1.50)	1.29 (1.32)	1.29 (1.32)	1.47 (1.51)	1.63 (1.68)	1.43 (1.47)	1.56 (1.60)
Reflections	18414 (1314)	24246 (1780)	13875 (772)	21923 (1615)	33797 (2337)	48474 (3576)	48915 (3563)	33813 (2446)	24034 (1662)	35537 (2423)	28172 (1930)
R <sub>merge</sub> (%)	7.0 (50.0)	6.1 (45.4)	7.7 (38.9)	7.4 (63.2)	9.3 (34.8)	8.0 (56.9)	6.7 (62.2)	7.1 (46.6)	6.4 (45.1)	7.6 (37.6)	10.6 (36.1)
Completeness (%)	99.7 (98.9)	99.8 (99.9)	80.0 (60.6)	99.2 (99.9)	99.5 (94.9)	98.8 (100.0)	99.9 (100.0)	99.8 (99.6)	97.0 (93.1)	99.4 (93.1)	99.3 (93.6)
<I>/<σ(I)>	23.2 (3.4)	22.8 (3.3)	11.8 (2.6)	15.2 (2.4)	13.8 (3.6)	21.0 (3.6)	26.9 (3.3)	24.3 (4.5)	14.5 (2.9)	17.3 (4.2)	14.6 (5.0)
R-factor (%)	18.7 (27.8)	19.1 (34.9)	19.6 (32.3)	19.1 (32.4)	18.0 (25.6)	17.8 (21.4)	17.2 (23.5)	18.1 (24.6)	20.8 (41.4)	18.1 (23.4)	18.3 (19.6)
R-free (%)	21.2 (30.2)	22.1 (44.3)	23.3 (34.3)	23.0 (38.9)	21.3 (32.3)	19.1 (21.6)	19.1 (28.6)	20.8 (28.6)	24.0 (55.0)	20.1 (27.1)	20.5 (25.6)
$\Delta_{\text{bond lengths}}$ (Å)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01
$\Delta_{\text{bond angles}}$ (°)	1.25	1.15	0.99	1.23	1.08	1.07	1.22	1.10	1.58	1.13	1.13
PDB code	2RAY	2RAZ	2RB0	2RB1	2RB2	2RBN	2RBO	2RBP	2RBQ	2RBR	2RBS
DOCK RMSD (Å)	0.82	1.44	0.60	0.62	1.42	1.29	2.04/1.12 <sup>b</sup>	0.97	1.44	1.16	1.93/1.84 <sup>b</sup>
AMBER RMSD (Å)	0.64	1.46	0.52	0.28	0.83	0.91	2.00/0.81 <sup>b</sup>	NA	0.87	0.93	NA
PLOP RMSD (Å)	0.54	1.49	0.58	0.37	1.08	1.10	2.00/0.65 <sup>b</sup>	0.63	1.61	1.02	1.80/1.70 <sup>b</sup>

All crystals belong to space group  $P3_22_1$

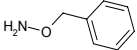
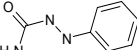
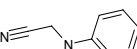
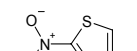
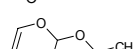
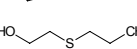
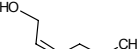
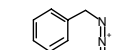
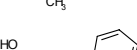
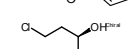
<sup>a</sup> Values in parentheses are for the highest resolution shell. <sup>b</sup> Two conformations of the crystallographic ligand were modeled.

**New L99A/M102Q Ligands Predicted by Rescoring.** Ten representative compounds that scored well by the MM-GBSA methods were experimentally tested for binding to the polar cavity (Table 3). These compounds were ranked poorly by docking, again typically because they were too large for the conformation of the cavity targeted by docking. Binding was detected at millimolar concentrations by  $T_m$  upshift for six of these ten compounds; for the remaining four binding was not observed (Table 3). We note, however, that for one of these four, 2-(n-propylthio) ethanol, we were able to determine a crystal structure in complex with the ligand by soaking a crystal of L99A/M102Q with 100 mM of compound, suggesting that it is a weak ligand for this cavity. AMBERDOCK correctly predicted binding for four of the six ligands that it suggested should bind, while incorrectly predicted binding for o-benzylhydroxylamine and 1-phenylsemicarbazide. Of the remaining two hits tested, prioritized by a high PLOP ranking, AMBERDOCK missed one real ligand but correctly distinguished one real decoy, ranking both compounds worse than 500. Two of the prospectively tested molecules were not rescored by AMBERDOCK because docking ranked them worse than 5000. PLOP correctly predicted binding for five of the six ligands that it suggested should bind but incorrectly predicted binding for cis-2-hexenol. Of the remaining hits tested, prioritized for testing by AMBERDOCK, PLOP missed two true ligands but correctly distinguished two decoys by ranking them worse than 1000.



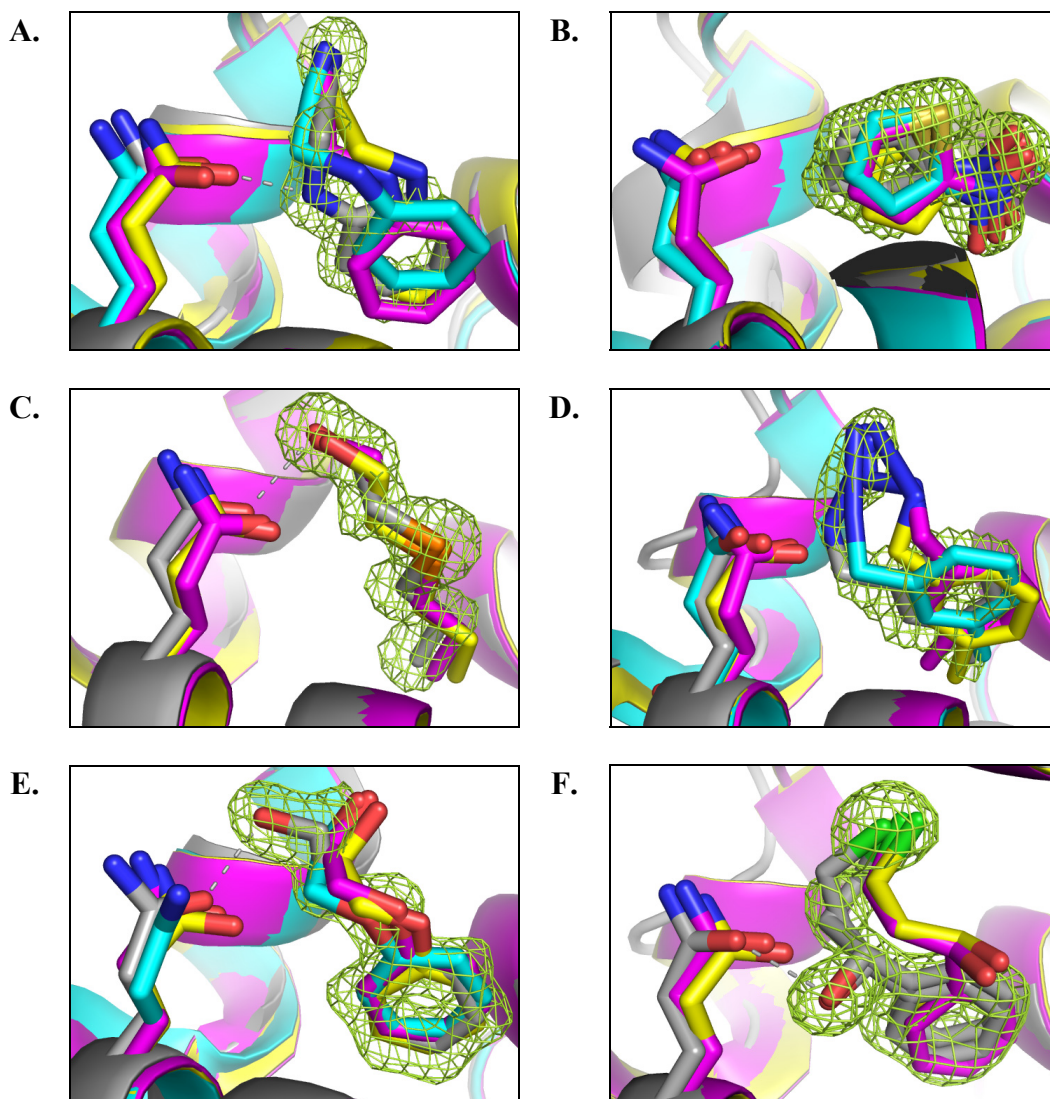
**Table 3.** Compounds predicted by AMBERDOCK and PLOP to bind to T4

Lysozyme L99A/M102Q.

Structure	Compound	Score and Rank <sup>a</sup>			C  <sup>b</sup> (mM)	pH	$\Delta T_m$ (°C)	Binding detected	Structure determined
		DOCK	AMBER	PLOP					
	o-Benzylhydroxylamine	-11.35 (647)	<b>-28.05</b> <b>(1)</b>	-14.14 (2271)	10	6.8	-0.6	No	NA
	1-Phenylsemicarbazide	-3.76 (3783)	<b>-26.79</b> <b>(4)</b>	-16.42 (1354)	10	6.8	0.0 <sup>c</sup>	No	NA
	N-Phenylglycinonitrile	-8.60 (1556)	<b>-25.47</b> <b>(11)</b>	<b>-40.17</b> <b>(11)*</b>	<10	3.0	5.1	Yes	Yes
	2-Nitrothiophene	-12.82 (318)	<b>-24.52</b> <b>(13)</b>	-16.94 (1165)	<10	3.0	4.4	Yes	Yes
	2-Ethoxy-3,4-dihydro-2H-pyran	-7.14 (2215)	<b>-24.21</b> <b>(14)</b>	-15.18 (1824)	10	3.0 6.8	1.3 -0.8	Weak	NA
	2-(N-Propylthio)ethanol	6.02 (6847)	NR <sup>d</sup>	<b>-27.20</b> <b>(20)</b>	10	3.0	0.1	No	Yes
	Cis-2-Hexen-1-ol	-1.58 (4291)	-10.25 (2260)	<b>-27.19</b> <b>(21)</b>	10	3.0	0.0	No	Yes
	3-Methylbenzylazide	-5.35 (2740)	<b>-20.51</b> <b>(116)</b>	<b>-25.87</b> <b>(35)</b>	10	3.0	1.9	Yes	Yes
	2-Phenoxyethanol	-4.08 (3270)	-16.53 (551)	<b>-25.82</b> <b>(36)</b>	10	3.0	1.2	Yes	Yes
	(R)(+)-3-Chloro-1-Phenyl-1-Propanol	3.6 (6074)	NR <sup>d</sup>	<b>-25.65</b> <b>(37)</b>	10	3.0	7.8	Yes	Yes

<sup>a</sup> Compound scores and ranks (in parenthesis) for DOCK, AMBERDOCK, and PLOP. Scores and ranks in bold font indicate ligands which rank in the top 200 for the respective scoring function. <sup>b</sup> Concentration at which ligand was tested. <sup>c</sup>  $\Delta T_m$  monitored using fluorescence at  $\lambda=291.5\text{nm}$  and measuring the integrated emission above 300 nm. <sup>d</sup> NR is not ranked.

Crystal structures of six L99A/M102Q ligand complexes were determined to compare predicted and experimental poses of these new ligands (Figure 4). Electron density for each ligand was unambiguous and was detailed enough to suggest two binding modes for 2-nitrothiophene and 3-chloro-1-phenyl-1-propanol. Docking predicted the pose of one ligand, 2-(n-propylthio)ethanol, to within 1 Å RMSD, while AMBERDOCK further minimized five of its six ligands and PLOP minimized three of its six ligands to within 1 Å RMSD (Table 2). Although the MM-GBSA methods collectively improved the binding mode predictions of all but one ligand, the key hydrogen bond interaction was missed in three of these structures (Figure 4a, e, and f). In addition, the azide group of 3-methylbenzylazide was incorrectly parameterized by both AMBERDOCK and PLOP, as was also observed in the L99A cavity. Neither DOCK nor the MM-GBSA rescoring correctly predicted the binding mode for 3-chloro-1-phenyl-1-propanol, with RMSD values of 1.9 and 1.7 Å, respectively. In three structures—2-nitrothiophene, 3-methylbenzylazide, and 3-chloro-1-phenyl-1-propanol—the F-helix of the cavity moves to accommodate the ligands while keeping the cavity still buried from solvent. In the complexes with 2-(n-propylthio)ethanol and 2-phenoxyethanol, there is evidence of a second conformation of residue Phe114 within the cavity that rotates and opens a water channel to the surface of the protein. Neither the helix movement nor the Phe114 rotation was sampled by the MM-GBSA methods.



**Figure 4.** Predicted and experimental ligand orientations for the polar L99A/M102Q cavity site. The carbons of the crystallographic, DOCK, AMBERDOCK, and PLOP predicted poses are colored grey, yellow, cyan, and magenta, respectively. Hydrogen bonds are depicted with dashed lines. The  $f_o-f_c$  electron density omit maps (green mesh) are contoured at 2.5-3.0 $\sigma$ . A. n-phenylglycinonitrile, B. 2-nitrothiophene, C. 2-(n-propylthio)ethanol, D. 3-methylbenzylazide, E. 2-phenoxyethanol, and F. (R)-(+)-3-chloro-1-phenyl-1-propanol bound to L99A/M102Q. Rendered with the program PyMOL.<sup>110</sup>

**New W191G ligands Predicted by Rescoring.** Fourteen representative compounds reprioritized to score well by the MM-GBSA methods but scored poorly by docking were experimentally tested for binding by measuring perturbation of the heme Soret band in CCP (Table 4).<sup>101</sup> Binding was detected for ten of these compounds at concentrations ranging from 50  $\mu$ M to 10 mM. Of the eleven compounds that AMBERDOCK predicted to bind with ranks better than 500, binding was detected for eight. Of the remaining prospective hits tested, AMBERDOCK correctly distinguished one compound as a decoy but missed two ligands by ranking them worse than 500. Of the nine compounds that PLOP predicted to bind with ranks better than 500, binding was detected for eight. Of the remaining prospective hits tested, PLOP missed two ligands but correctly distinguished three decoys, ranking them worse than 500.

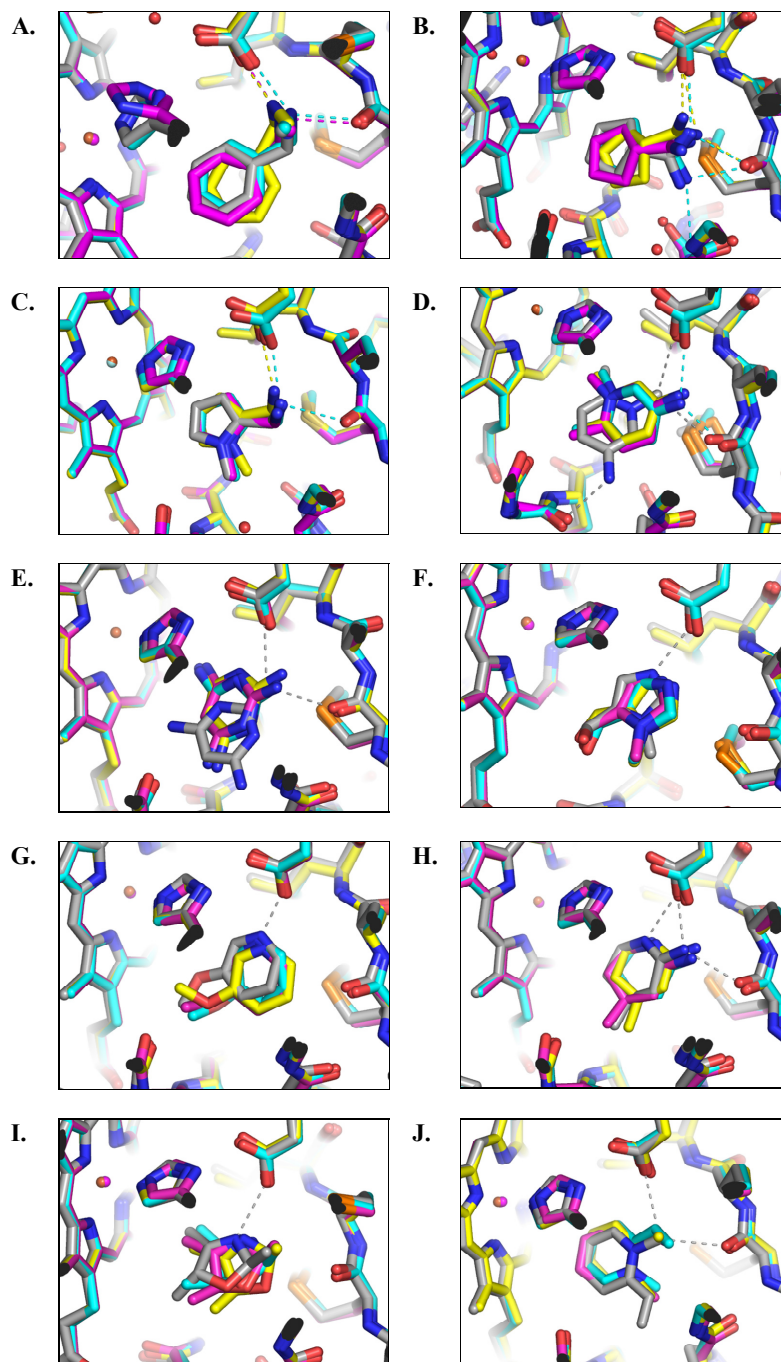
**Table 4.** Compounds predicted to bind by AMBERDOCK and PLOP to CCP

W191G.

Structure	Compound	Score and Rank <sup>a</sup>			C  <sup>b</sup> (mM)	Binding detected	Structure determined
		DOCK	AMBER	PLOP			
	N-Methyl-1,2-Phenylene Diamine	-20.6 (618)	<b>347.03</b> <b>(30)</b>	-38.08 (530)	10.0	No	N/A
	N-Methylbenzyl-Amine	-18.59 (942)	<b>347.85</b> <b>(38)</b>	-16.21 (952)	1.0	Yes	Yes
	Cyclopentane-Carboximidamide	-13.38 (2134)	<b>347.86</b> <b>(39)</b>	-44.39 (389)	1.0	Yes	Yes
	(1-Methyl-1H-Pyrrol-2-yl)-Methylamine	-14.74 (1830)	<b>348.17</b> <b>(49)</b>	-31.88 (796)	0.05	Yes	Yes
	5-Nitro-6-Aminouracil	-12.14 (2435)	<b>348.49</b> <b>(62)</b>	-31.47 (827)	1.0	No	N/A
	1,2-Dimethyl-1H-Pyridin-5-Amine	-22.95 (362)	<b>349.34</b> <b>(87)</b>	<b>-54.67</b> <b>(59)</b>	0.5	Yes	Yes
	2-Aminobenzylamine	-12.62 (2316)	<b>349.34</b> <b>(96)</b>	-34.19 (671)	10.0	No	N/A
	Pyrimidine-2,4,6-Triamine	<b>-36.54</b> <b>(7)</b>	<b>344.29</b> <b>(12)</b>	<b>-59.87</b> <b>(53)</b>	1.0	Yes	Yes
	1,3-Dimethyl-2-Oxo-2,3-Dihydro-Pyrimidin-1-ium	-8.52 (3093)	363.47 (1901)	<b>-56.65</b> <b>(32)</b>	10.0	No	N/A
	1-Methyl-5-Imidazolecarboxaldehyde	-21.14 (551)	358.53 (746)	<b>-57.12</b> <b>(28)</b>	10.0	Yes	Yes
	3-Methoxypyridine	-23.05 (2665)	355.17 (393)	<b>-55.31</b> <b>(44)</b>	10.0	Yes	Yes
	2-Imino-4-Methylpiperdine	-17.30 (1695)	<b>349.17</b> <b>(82)</b>	<b>-52.43</b> <b>(119)</b>	10.0	Yes	Yes
	2,4,5-Trimethyl-3-Oxazoline	-13.96 (1962)	355.98 (455)	<b>-52.32</b> <b>(124)</b>	0.25	Yes	Yes
	1-Methyl-2-Vinyl-Pyridinium	-15.17 (1716)	363.60 (1938)	<b>-52.32</b> <b>(125)</b>	0.5	Yes	Yes

<sup>a</sup> Compound scores and ranks (in parenthesis) for DOCK, AMBERDOCK, and PLOP. Scores and ranks in bold font indicate ligands which rank in the top 200 for the respective scoring function. <sup>b</sup> Concentration at which ligand was tested.

Crystal structures of CCP in complex with the ten new ligands were obtained (Figure 5). The electron density for the ligands was unambiguous. Docking predicted three structures to within 1 Å of the crystallographic result whereas the MM-GBSA methods did so for seven structures, typically with improved hydrogen bonding interactions (Table 5). For three ligands, the docking poses were over 1.9 Å away from the crystallographic results, and MM-GBSA refinement did little to improve these structures. In four of the complex structures—cyclopentane-carboximidamide, 1,2-dimethyl-1H-pyridine-5-amine, 1-methyl-2-vinyl-pyridinium, and pyrimidine-2,4,6-triamine—the loop composed of residues 190-195 flips out by nearly 12 Å opening the cavity to bulk solvent. This large loop motion was not sampled by MM-GBSA.



**Figure 5.** Predicted and experimental ligand orientations for the anionic CCP cavity. The carbons of the crystallographic, the DOCK, AMBERDOCK, and PLOP predicted poses are colored grey, green, cyan, and orange, respectively. A. n-methylbenzylamine, B. cyclopentane carboximidamide, C. (1-methyl-1H-pyrrol-2-yl)-methylamine, D. 1,2-dimethyl-1H-pyridin-5-amine, E. pyrimidine-2,4,6-triamine, F. 1-methyl-5-imidazolecarboxaldehyde, G. 3-methoxypyridine, H. 2-imino-4-methylpiperidine, I. 2,4,5-trimethyl-3-oxazoline, and J. 1-methyl-2-vinylpyridinium. Rendered with the program PyMOL.<sup>110</sup>

**Table 5.** Crystallographic data and the RMSD values for predicted and crystallographic ligand geometries in the CCP site.

	n-methyl Benzylamine	cyclopentane carbox imidamide	(1-methyl-1H- pyrrol-2-yl)- methylamine	1,2-dimethyl- 1H-pyridin-5- amine	pyrimidine- 2,4,6- triamine	1-methyl-5- imidazole carbox aldehyde	3-methoxy pyridine	2-imino-4- methyl piperidine	2,4,5- trimethyl-3- oxazoline	1-methyl-2- vinyl- pyridinium
Resolution (Å)	1.24 (1.27)	1.80 (1.85)	1.39 (1.43)	1.50 (1.54)	1.50 (1.54)	1.50 (1.54)	1.80 (1.85)	1.50 (1.54)	2.49 (2.56)	1.50 (1.54)
Reflections	104081 (4744)	36504 (2670)	78561 (5188)	57782 (4371)	63009 (4512)	63445 (4489)	36874 (2587)	63783 (4467)	11042 (680)	63108 (3364)
R <sub>merge</sub> (%)	6.1 (40.7)	5.3 (10.9)	4.5 (20.2)	5.1 (22.4)	5.2 (38.5)	3.8 (22.1)	6.4 (23.3)	5.0 (33.4)	2.7 (5.9)	4.2 (19.6)
Completeness (%)	93.1 (58.0)	99.4 (99.9)	99.0 (90.5)	92.0 (95.9)	99.8 (97.9)	99.7 (96.3)	99.5 (96.2)	99.7 (96.5)	82.3 (69.7)	99.3 (95.4)
<I>/<σ(I)>	29.9 (1.8)	51.7 (30.1)	43.2 (5.6)	41.7 (6.0)	30.6 (2.8)	37.8 (5.2)	34.8 (9.2)	31.0 (3.5)	26.1 (12.9)	39.2 (5.9)
R-factor (%)	12.2 (24.8)	15.9 (19.3)	13.8 (20.7)	14.4 (17.8)	15.2 (21.0)	14.5 (16.9)	15.0 (18.4)	14.3 (17.9)	17.7 (21.9)	14.6 (16.3)
R-free (%)	14.6 (24.4)	19.5 (24.0)	15.4 (24.1)	16.7 (25.5)	17.3 (25.7)	16.4 (21.0)	19.1 (24.0)	16.9 (23.9)	22.9 (29.4)	16.7 (21.7)
Δ <sub>bond</sub> lengths (Å)	0.01	0.02	0.01	0.01	0.01	0.01	0.02	0.01	0.02	0.01
Δ <sub>bond</sub> angles (°)	1.59	1.39	1.19	1.22	1.14	1.24	1.50	1.22	1.95	1.14
PDB code	2RBT	2RBU	2RBV	2RBW	2RBX	2RBY	2RBZ	2RC0	2RC1	2RC2
DOCK RMSD (Å)	1.05	1.34	0.58	2.97/3.16 <sup>b</sup>	1.91	1.07	0.89	0.55	1.32	2.58
AMBER RMSD (Å)	0.37	0.31	0.34	2.84/3.01 <sup>b</sup>	1.70	1.06	0.81	0.50	0.83	2.55
PLOP RMSD (Å)	0.45	0.99	0.30	2.98/3.15 <sup>b</sup>	1.91	0.94	0.77	0.33	0.81	2.59

All crystals belong to space group  $P2_12_12_1$

<sup>a</sup> Values in parentheses are for the highest resolution shell. <sup>b</sup> Two conformations of the crystallographic ligand were modeled.



**Overall Performance in Predicting Top 100 Hits.** The simplicity of these model cavity sites, the number of known ligands and decoys, and our experience testing their ligands<sup>59, 65, 100, 103, 104</sup> often allow us to predict what will turn out to be true ligands and true decoys from among top-scoring molecules, based on their physical properties. We examined the top 100 hits predicted to bind by docking and MM-GBSA, compared property distributions, and made educated guesses as to whether or not they will bind. The 100 top ranking MM-GBSA rescored compounds for the L99A and L99A/M102Q cavities were larger, more flexible, and more polar, with more hydrogen bond acceptors and lower ClogP values per heavy atom compared to the top 100 hits from docking. For the anionic W191G cavity there was a similar trend towards larger molecules and also a drift away from the singly charged cations favored by DOCK, with more dications and neutral molecules prioritized among the top ranking 100 molecules by the MM-GBSA methods. The increased size and greater differences in polarity of the molecules in the MM-GBSA hit lists resulted in lower mean pair-wise similarities among the molecules, and consequently, an increase in the diversity of the rescored hit lists relative to the docking hit lists. Thus, using ECFP\_4 fingerprints (SciTegic, Inc.), the average pair-wise Tanimoto coefficient among the 100 top docking molecules for the L99A cavity with DOCK, AMBERDOCK, and PLOP was 0.17, 0.12, and 0.10, respectively. Similar trends were observed in the other two cavities. The same tendencies that led to greater diversity in ligands and their properties, however, reduced the raw hit rates we anticipate from among the top 100 ranking MM-GBSA ligands compared to those predicted by docking (Table 6). For example, among the top 100 docking hits for the CCP cavity there were 29 true ligands and no experimentally determined decoys. Of the remaining molecules—all untested—were what we predict to be 79 likely ligands and 7 likely decoys, based on their similarity to known ligands and decoys and their physical properties such as size and charge complementarity. Conversely, among the top 100 PLOP hits for the anionic cavity were only 15 experimentally tested ligands and 1 experimental decoy. Among the

untested molecules were what we suspect are 53 further ligands and 22 further decoys. Among the top AMBERDOCK hits for this cavity were 19 true ligands and 3 experimental decoys. Among the untested molecules prioritized by this program, we suspect that there are 67 further ligands and 14 more decoys. Similar trends were observed in the other two cavities (Table 6). Admittedly, these numbers reflect guesses only, but we suspect that the overall trends would be born out by experiment (the interested reader may draw their own conclusions from the full lists in Supplementary Materials Tables 1-9) Thus, whereas the MM-GBSA methods rescued many docking false negatives and sampled a more diverse chemical space among the top hits, they also suggested more false positives among the very top-scoring molecules and, we suspect, have a lower overall hit-rate in this segment of the molecules prioritized for testing.

**Table 6.** Likely ligands and decoys among the top 100 ranked ligands by docking and MM-GBSA.

Method	True ligands in top hits	True decoys in top hits	Likely ligands in top 100 hits <sup>a</sup>	Likely decoys in top 100 hits <sup>b</sup>	Ambiguous <sup>c</sup>
<b>L99A cavity</b>					
DOCK	7	3	63	23	14
PLOP	6	1	35	22	43
AMBERDOCK	8	2	54	25	21
<b>L99A/M102Q cavity</b>					
DOCK	13	2	73	12	15
PLOP	5	1	31	8	61
AMBERDOCK	7	2	43	22	35
<b>W191G cavity</b>					
DOCK	29	0	79	7	14
PLOP	15	1	53	22	25
AMBERDOCK	19	3	67	14	19

<sup>a</sup> Molecules that, based on their physical properties and similarity to known ligands, are likely to be cavity ligands. <sup>b</sup> Molecules that, based on their physical properties and similarity to known decoys, are likely not to bind. <sup>c</sup> Molecules that are sufficiently different from known ligands and decoys, and whose physical properties are not sufficiently distinctive, such that no prediction was made (for L99A and L99A/M102Q molecules). For W191G, molecules that were mis-protonated during database preparation, relative to the expected protonation at pH 4.5, and so are not counted to measure the performance of the scoring function.

**Origins of False Positive Hits Suggested by MM-GBSA Rescoring.** In these simple cavities, false-positive hits often identify specific pathologies in a scoring function. For example, the MM-GBSA methods seemed distracted by compounds bearing what is almost certainly the wrong net charge for the W191G cavity, which extensive testing has shown preferentially binds mono-cations over neutral molecules (few of which have been observed to bind, and then only weakly) and dications (none of which have been observed to bind). For instance, among the top 100 ranking molecules predicted by PLOP, there were 13 dications. Whereas AMBERDOCK predicted only one dication, it prioritized five neutral molecules among the top 100 hits. The dications will pay too high a desolvation penalty to be compensated by the interaction with the single anion in the site (Asp235), and the neutral compounds desolvate the same aspartate without recouping enough in interaction energy. Balancing polar and ionic interactions with concomitant solvation penalties is a challenge for the field, one clearly faced by these methods as well. On the other hand, many of the top ranked PLOP ligands for L99A (47 out of the top 50) and L99A/M102Q contained one or more nitriles. While some of these compounds may well be ligands, as in the case of n-phenylglycinonitrile for L99A (Table 1), we suspect that this represents a ligand parameterization pathology in accounting for the distribution of atomic partial charge of the nitrile group as opposed to a genuinely meaningful enrichment (we excluded these compounds from the PLOP rescored hit lists for L99A and L99A/M102Q; they do not contribute to the accounting described in this work). This highlights the importance of good ligand parameterization for database screening—which is a considerable challenge for hundreds-of-thousands of molecules typically screened by docking—the lack of which can undermine any improvement in theory.

## Discussion

In principle, the most important improvements of MM-GBSA over docking, certainly over the program used in this study, DOCK3.5.54, are the better representation of electrostatic interactions, ligand and protein desolvation energies, and relaxation of the ligand-protein complex. The simplicity of the model cavity sites allows us to explore how these terms influence docking results in detail and to make prospective predictions for ligands that we can, in fact, acquire and test. Many investigators will be unsurprised to see that the MM-GBSA methods can rescue molecules that rank poorly in the docking calculation owing to the rigid-receptor approximation used in docking. Ligands that were too big to be accommodated well in the original docking are well-fit by a binding site that has been allowed to relax by energy minimization and, in the case of AMBERDOCK, short MD simulations. This was true both in retrospective calculations as well as in prospective predictions. The ability to relax the site also resulted in rescored hit lists that were more diverse with a wider range of likely ligands. Perhaps less anticipated was the cost of allowing such conformational change—some of the rescued, high-scoring molecules by MM-GBSA do not, in fact, bind to the cavity sites. These molecules are new false-positives introduced by the higher level of theory. Indeed, the overall hit rates at the very top of the ranked lists are arguably better by simple docking than by MM-GBSA rescoring, at least when evaluated simplistically by the raw number of hits and likely hits (this is arguably offset by the greater diversity of the MM-GBSA hit lists). Partly this reflects problems in ligand parameterization, and partly difficulties in the treatment of the electrostatics in the binding sites. The most important challenge for MM-GBSA and for flexible receptor models in general is balancing the opportunities to find new ligands as receptor geometries are relaxed with the introduction of new false positives as the need to consider large receptor internal energies is introduced. Specific

examples of these opportunities and problems are apparent in the three cavity sites studied here.

The principal improvement conferred by MM-GBSA rescoring in the model cavity sites over docking was the inclusion of receptor binding site relaxation, which improved the ranks of larger ligands that rigid receptor docking missed. AMBERDOCK, for example, correctly predicted 2-ethoxyphenol to bind to L99A (Table 1, Figure 3d). This compound is too large for the unrelaxed conformation of this cavity targeted by docking, but minimization and MD simulations allow the ligand to be well accommodated by effectively expanding the site. Often, this relaxation led not only to improved rankings but also improved geometries. For many ligands, RMSD values between the MM-GBSA predictions and the crystallographic results declined relative to those of the docking predictions and, especially in the W191G anionic cavity, many ligands refined by MM-GBSA had improved hydrogen bonding to the site. Examples of this include the new W191G cavity ligands *n*-methylbenzylamine and cyclopentane-carboximidamide (Table 4, Figures 5a and 5b, respectively).

The structural relaxation with MM-GBSA performed well when the initial docking geometry resembled the crystallographic pose, but did little when large protein conformational changes were provoked by ligand binding. For instance, F-helix unwinding and rotamer change by Val111 in L99A and L99A/M102Q were never captured by the method, nor was the extensive loop flipping observed in several of the W191G-ligand complexes. When such movements occurred, MM-GBSA rescoring could not rescue substantially incorrect docking poses, such as that adopted by 3-chloro-1-phenyl-propanol for L99A/M102Q (Table 3, Figure 4f) and pyrimidine-2,4,6-triamine predicted for CCP (Table 4, Figure 5e), notwithstanding the large improvement in their rankings conferred by the rescoring. These large movements are outside the radius of convergence of the local relaxation undertaken by the MM-GBSA methods. Indeed, even more time-consuming thermodynamic integration methods are hard put to sample such

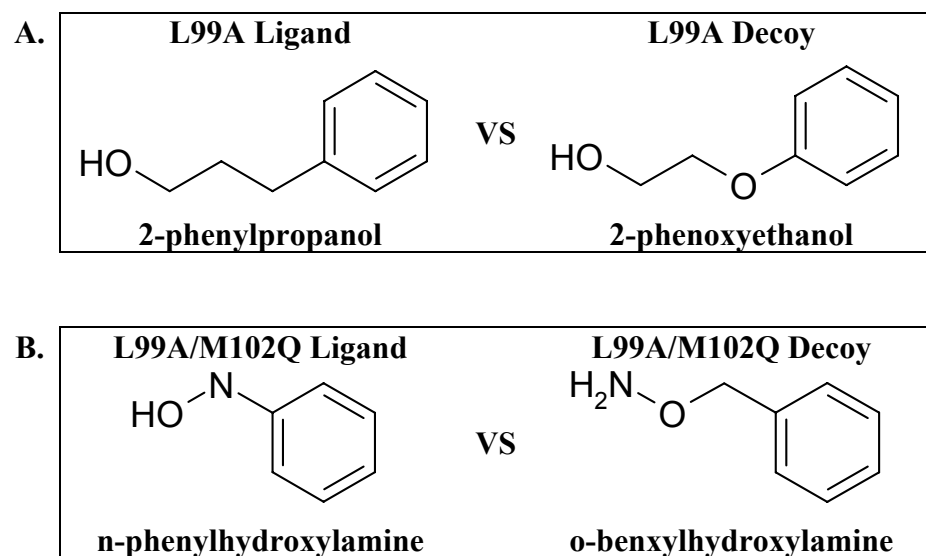
changes without explicit “confine-and-release” strategies, which depend on a foreknowledge that such movements are likely.<sup>114</sup> And whereas loop sampling methods have had encouraging successes in predicting such large movements,<sup>115</sup> this remains a frontier challenge for ligand and structure prediction methods.

Pragmatically, the inability to predict the structural accommodations provoked by some large ligands is offset by the correct re-prioritization of what were docking false-negatives as ligands. The same comfort is not afforded by the ten false negatives introduced by the MM-GBSA methods, nor by the lower overall hit rates compared to docking among the very top scoring ligands (Table 6). By allowing the receptor to respond to ligand binding, one allows for new and potentially unfavorable receptor conformations. These must be distinguished by the MM-GBSA energy functions from the true low-energy conformations that may be sampled in solution. This is challenging as the receptor conformational energies are large, and the errors in these calculations are typically on the same order of the net interaction energy of the protein-ligand complex. Although some of the errors are cancelled by subtraction of the internal energies before and after ligand binding, one is still subtracting two large numbers with relatively large errors to find a small one, the net binding free energy. Consistent with this view, ligands achieved their maximal advantage over decoys on rescoring when we allowed only a 5 Å region around the binding site to relax. Allowing the full protein to relax, or even an 8 Å region around the binding site, diminished the discrimination of known ligands from decoys. Of course, relaxing the entire system is the more physically correct way to calculate these energies. Falling back on limited relaxation speaks to a larger methodological issue.

The three cavity sites targeted here are contrivances of human design and ligands discovered for them have no intrinsic value other than for testing methods. Indeed, in these simple model systems the failures are often more interesting than the successes, as they can illuminate a specific methodological problem.<sup>59, 65, 100, 103, 104</sup> Examples are the

ten false positives predicted for the cavity sites by MM-GBSA rescoring. Some of these reflect ligand parameterization problems. For instance, the highly polar 1-phenylsemicarbazide was predicted to bind to both L99A and L99A/M102Q by AMBERDOCK (Table 1 and 3), presumably reflecting a failure to properly parameterize the partial charges of the semi-carbazide functionality and so account for its high desolvation energy. Similarly, we suspect that the many nitrile containing decoys predicted by PLOP for L99A and L99A/M102Q are examples of failures in ligand parameterization. These are in some senses trivial failures that may be addressed by close attention to particular ligand groups and improved partial atomic charge models; admittedly, this can be a daunting task for screening databases containing hundreds of thousands of disparate molecules. More interesting are the eight false-positives that are true energy function decoys. Several of these highlight difficulties in the treatment of electrostatics and solvation in the binding sites. 2-phenoxyethanol, for example, was predicted to bind by both PLOP and AMBERDOCK to L99A (Table 3). This decoy has a similar topology to 2-phenylpropanol, a known ligand<sup>63</sup> (Figure 6a); however, the ether of 2-phenoxyethanol increases its polarity and presumably its solvation energy, which is not fully captured by the MM-GBSA implicit solvent model. Similarly, o-benzylhydroxylamine was the top-ranking AMBERDOCK hit for L99A/M102Q, but is a decoy (Table 3). The terminal -ONH<sub>2</sub> of this compound is too polar for the site, stranding one unpaired polar hydrogen from the NH<sub>2</sub> group in this largely hydrophobic site. Interestingly, the polar cavity does bind n-phenylhydroxylamine (unpublished data), which has the same hydrogen bond accounting as o-benzylhydroxylamine and topologically resembles it closely (Figure 6b). The difference between these nearly identical molecules is that in the former the two hydrogen bond donors from the ligand can both be accommodated by the carbonyl of the receptor glutamine, whereas in the decoy both hydrogen bond donors originate from the same atom—the nitrogen of the o-benzylhydroxylamine—and only one can be accommodated by the carbonyl oxygen.





**Figure 6.** Topologically similar ligands and decoys. A. 2-phenylpropanol and 2-phenoxyethanol for L99A and B. n-phenylhydroxylamine and o-benzyloxyamine to L99A/M102Q.

The challenges of balancing ligand electrostatic interaction energies and desolvation penalties were also apparent in the anionic, W191G cavity. Most obvious were those molecules that did not bear the correct mono-cationic charge state. The 13 molecules that were doubly charged among the top scoring PLOP hits are almost certainly decoys, and this is also the case for the AMBERDOCK false positive 5-nitro-6-aminouracil, which is neutral and cannot make the ion-pair interaction with Asp235 (Table 4). More subtly, whereas 1,3-dimethyl-2-oxo-2,3-dihydropyrimidin-1-ium is charged, this charge is shared between the two cyclic nitrogens and results in a compound with reduced electrophilicity compared to a compound with a localized charge. The AMBERDOCK false-positives n-methyl-1,2-phenylene-diamine and 2-aminobenzylamine (Table 4) most likely do not bind because of steric clashes that inhibit optimal positioning of the charge-charge interaction. These failures point to specific directions for improved treatment of the balance between electrostatic interaction and desolvation energies in the MM-GBSA methods.

Overall, the results of MM-GBSA rescoring of docking hit lists on the model binding sites seem conflicted. On the one hand, rescoring rescued many docking false negatives, improved the geometric fidelity of most of the predicted structures, and increased the diversity of the hit lists. On the other hand, rescoring introduced more false-positives, especially among the very top ranking ligands, compared to the simpler docking protocol. These observations may be reconciled by recognizing that what is probably the greatest advantage of the MM-GBSA methods over docking for the model sites, the relaxation of the protein-ligand complex, also presents the greatest challenge to discrimination. To allow a flexible receptor, one must consider the relative energies of the different protein conformations explored. This implicates the pair-wise interactions of thousands of protein atoms, as opposed to the tens of atoms involved in the immediate protein-ligand complex. To properly rank the energies of the complexes, one must also properly account for the larger uncertainties that accompany the much higher magnitude

energies of the overall system. Whereas this is the thermodynamically correct approach, it introduces many interactions that have little bearing on the intimacies of the protein-ligand complex itself. Rigid receptor docking, for all the calumny poured upon it, can ignore these large magnitude yet low relevance interactions. Of course, this leads to many false-negatives, but it avoids many of the false positives to which the MM-GBSA methods are prone. Pragmatically, this suggests that hits derived from docking to a rigid experimental receptor conformation—and ideally more than one<sup>65, 82</sup>—and hits prioritized by rescoring after MM-GBSA refinement with binding site minimization will provide good candidates for experimental testing. Despite its greater sophistication, MM-GBSA rescoring has a harder task, and its predictions will not, by every criterion, be better than those of a modern docking program; rather, our results suggest they will complement and add to them. More generally, however, the MM-GBSA and related methods with binding site relaxation do represent a higher level of theory; extending them and related methods seems ultimately necessary for fundamental improvements in molecular docking and structure-based screening, which are so actively sought.<sup>116</sup>

## Materials and Methods

**Docking Against Cavity Sites.** DOCK3.5.54<sup>59, 67</sup> was used to dock a multi-conformer database of small molecules into the model cavity sites. The receptors, grids, spheres, and ligand databases were prepared as described for the T4 Lysozyme<sup>59</sup> and CCP<sup>103</sup> cavities, respectively. Briefly, to sample ligand orientations, ligand, receptor, and overlap bins were set to 0.2 Å; the distance tolerance for matching ligand atoms to receptor was set to 0.75 Å. Each docking pose was evaluated for steric fit. Compounds passing this filter were scored for electrostatic and van der Waals complementarity and assigned the full penalty for transfer from a dielectric of 80 to one of 2, as calculated by AMSOL.<sup>71, 72</sup> Sampling and scoring required less than a second per ligand on a single 3.2 GHz Xeon processor. The best scoring conformation of each of the 10,000 top scoring molecules against L99A and L99A/M102Q and the 5400 top scoring molecules against CCP were saved and rescored by the MM-GBSA protocols.

**Rescoring with PLOP.** The rescoring procedure with Protein Local Optimization Program (PLOP)<sup>111, 112</sup> was essentially as described.<sup>96</sup> Ligand parameters were calculated with IMPACT.<sup>117</sup> The partial atomic charges of the ligands were replaced by the AM1-CM2 charges calculated by AMSOL (v6.5.3) as these were the same charges used during the initial docking.<sup>59</sup> The same protein structure file used in docking was used for rescoring. Protein parameters were defined by IMPACT with the exception of the partial charges for the heme cofactor in CCP W191G, which were the same as used in the docking method.<sup>103</sup> All energy minimizations were performed using PLOP with the all-atom OPLS force field (OPLS-AA)<sup>118</sup> and the Surface Generalized Born (SGB) implicit solvent model.<sup>119</sup> PLOP implements a multiscale truncated-Newton (MSTN) minimization algorithm as described.<sup>120</sup> For receptor minimization and calculation of  $E_{complex}$  and  $E_{receptor}$ , a pre-specified list of residues within 5 Å of the binding site were

minimized after an initial sidechain rotamer search. (Residues 78, 84, 85, 87, 88, 91, 98-100, 102, 103, 106, 111, 118, 121, 133, and 153 for L99A and L99A/M102Q and residues 174-180, 189-192, 202, 230-232, 235, and water 308 for CCP). The rotamer search algorithm is as described in supplementary materials.

Preliminary PLOP calculations of the hydrophobic and polar cavities were performed with a rigid receptor and resulted in very little separation of ligands and known decoys. On the other hand, PLOP calculations in which a larger set of residues (those within 8 Å of the binding site) were minimized and resulted in worse overall enrichments of known ligands and a decreased separation of known ligands and decoys relative to minimizing a smaller 5 Å pocket. To approximate a fully desolvated ligand and cavity for the hydrophobic L99A and polar L99A/M102Q sites, only the SGB solvation term of the free ligand was included in the calculation of the total PLOP binding energies. Initial PLOP calculations including the SGB solvation terms for the calculation of the complex and free protein energies resulted in poor enrichments of known ligands, decreased separation of ligands and known decoys, as well as an enrichment of hits with increased polarity and electrostatic interactions. For the more solvated CCP cavity, the SGB terms were included in the calculation of the complex, free protein, and free ligand energies for the total binding energy.

**Rescoring with AMBERDOCK.** AMBERDOCK is based on the `amber_score()` scoring module in DOCK6. The ligand structures were modified using the *antechamber* suite of programs to create input files that could be read by Leap to generate the parameter and topology files for AMBERDOCK. Antechamber<sup>121</sup> has been developed to be used with the general AMBER force field (GAFF) for small molecules.<sup>122</sup> Charges for the ligands were generated using three charge methods in Antechamber—PEOE,<sup>123</sup> AM1-BCC,<sup>124</sup> and HF/6-31G\* RESP.<sup>125</sup> The protonation states of the ligands were kept the same as the previous docking run for consistency in rescoring. AMBER *ff94*

parameters were assigned to all the protein atoms. The standard parameters for the heme cofactor as implemented in the Amber 9 program was used for the CCP cavity.<sup>126</sup> The protonation states of Histidine residues were predicted based on their close neighbors. The GB model corresponding to igb=5 in the AMBER 9 program was used.<sup>127</sup> The surface area term was calculated using the LCPO model.<sup>128</sup> A non-bonded cutoff of 18 Å was used for the calculations.

The starting structures were taken from the docked pose. The structures were subjected to 100 steps of conjugate gradient minimization, 3000 steps of MD simulation with a 1 fs time step at a temperature of 300K, followed by 100 steps of minimization. During the minimization and MD, only the ligand and the protein residues within 5 Å of the ligand were allowed to move. To expedite the scoring process, we calculated the energy of the free receptor ( $E_{receptor}$ ) once, and used this energy as a constant term during the subsequent energy evaluations for the rest of the ligands in the database. Binding free energy calculations with AMBERDOCK follows a scheme as described in supplementary materials. Several AMBERDOCK rescoring protocols with slight variations were retrospectively tested and results are described in supplementary materials.

**Protein Preparation and Expression.** T4 Lysozyme mutants L99A and L99A/M102Q and CCP mutant W191G were expressed and purified as described.<sup>59, 101</sup>

**Binding Detection of Ligands to T4 Lysozyme Cavities by Upshift of Thermal Denaturation Temperature.** To detect binding, L99A and L99A/M102Q were denatured reversibly by temperature in the presence and absence of the putative ligand. Molecules that bind preferentially to the folded cavity-containing protein should stabilize it relative to the *apo* protein, raising its temperature of melting.<sup>63</sup> All thermal melts were conducted in a Jasco J-715 spectropolarimeter as described.<sup>63</sup> Each compound was screened in its neutral form. All compounds tested against L99A and L99A/M102Q were

assayed in a pH 3 buffer containing 25 mM KCl, 2.9 mM phosphoric acid, and 17 mM  $\text{KH}_2\text{PO}_4$  with the exception of 1-phenylsemicarbazide and o-benzylhydroxylamine. To maintain compound neutrality, these two were assayed at pH 6.8 in a 50 mM potassium chloride and 38% (v/v) ethylene glycol buffer.<sup>63</sup> Thermal melts were monitored by far UV circular dichroism, except for melts in the presence of 4-(methylthio)nitrobenzene, 1-phenylsemicarbazide, and 2,6-difluorobenzylbromide, which absorb strongly in the far UV region. For these three, thermal denaturation was measured by the intensity of the integrated fluorescence emission for all wavelengths above 300 nm, exciting at 283 to 292 nm, using a fluorescence PMT on the Jasco instrument. Thermal melts were performed at a temperature ramp rate of 2 K/min. A least-squares fit of the two-state transition model was performed with the program EXAM<sup>129</sup> to calculate  $T_m$  and van't Hoff  $\Delta H$  values for the thermal denaturations. The  $\Delta C_p$  was set to 8  $\text{KJ mol}^{-1} \text{K}^{-1}$  (1.94  $\text{kcal mol}^{-1} \text{K}^{-1}$ ).

**Binding Detection of Ligands to CCP W191G.** Ligand binding was measured in 50 mM acetate buffer pH 4.5. To avoid competition in ligand binding with small cations like potassium,<sup>101</sup> the pH of the buffer was adjusted with Bis-Tris propane. The compounds were dissolved in dimethyl sulfoxide (DMSO). Binding of compounds to CCP was monitored by the red shift and increase of absorbance of the heme Soret band<sup>101</sup> at 10 °C.

**Structure Determination.** Crystals for L99A and L99A/M102Q were grown as described<sup>59</sup> and the resulting crystals belonged to space group  $P3_22_1$ . Crystals were soaked overnight to one week in crystallization buffer containing as much as 100 mM compound. In addition to soaking, drops of neat compound were added to the cover slip surrounding the drop containing the crystal. After soaking, the crystals were cryoprotected with a 50:50 Paraton-N (Hampton Research, Aliso Viejo, CA), mineral oil

mix. Crystals for CCP W191G were grown as described<sup>102</sup> and the resulting crystals belonged to space group  $P2_12_12_1$ . Crystals were soaked in 25% MPD with 1 to 50 mM compound for 4 hours or overnight with the exception of pyrimidine-2,4,6-triamine, which was soaked for 15 minutes.

Diffraction data for the complexes of L99A with  $\beta$ -chlorophenetole, 4-(methylthio)nitrobenzene, 2,6-difluorobenzylbromide and the complex of L99A/M102Q with 3-methylbenzylazide were collected using a Rigaku X-ray generator equipped with a rotating copper anode and a Raxis IV image plate. Data for the complexes of L99A/M102Q with *n*-phenylglycinonitrile and 2-nitrothiophene and the complex of CCP with *n*-methylbenzylamine were collected on Beamline 9-1 at the Stanford Synchrotron Radiation Laboratory (SSRL) using an ADSC-CDD detector. Data for all other complexes were collected on Beamline 8.3.1 of the Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory using an ADSC-CCD detector. All data sets were collected at 100 K. Reflections were indexed, integrated, and scaled using HKL2000.<sup>77</sup> Parameters for ligands were generated with PRODRG.<sup>130</sup> Complexes were refined using the CCP4 software package.<sup>131</sup> Interactive model building was performed using Coot.<sup>132</sup>

**Protein Data Bank Accession Codes.** The crystallographic coordinates for the complex structures presented in this work have been deposited with the RCSB Protein Data Bank with accession codes 2RAY, 2RAZ, 2RB0, 2RB1, 2RB2, 2RBN, 2RBO, 2RBP, 2RBQ, 2RBR, 2RBS, 2RBT, 2RBU, 2RBV, 2RBW, 2RBX, 2RBY, 2RBZ, 2RC0, 2RC1, and 2RC2.



## **Acknowledgments**

This work supported by GM59957 (to BKS), AI035707 (to MPJ) and GM56531 (to DAC). MPJ is a consultant to Schrodinger Inc. We thank Niu Huang for advice using PLOP; Michael Keiser and Jerome Hert for help with chemical similarity calculations; Michael Mysinger, Veena Thomas and Michael Keiser for reading this manuscript; and MDL for providing the ACD database.

### **Gloss to Chapter 3.**

A central challenge for structure-based drug design is to predict the binding free energy for a ligand to its protein target. The following chapter addresses how well alchemical free energy calculations can accomplish this feat for a simple hydrophobic cavity. This T4 lysozyme site was discussed in the previous two chapters. At its heart, the following chapter suggests a rescoring method similar to that of Chapter 2. Docking was first used to generate the starting geometries of ligands, which were then rescored by a more energetically sophisticated method. Chapter 2 described a “bottom-up” approach to rescoring that added only a few additional terms to docking. However, Chapter 3 describes a “top down” rescoring approach. First, we rescored using a very high level of theory that makes fewer approximations than docking or the MM-GBSA methods described in Chapter 2. Then, we tried breaking the method by taking away energetic sophistication and adding in more approximations.

The free energy methods discussed in this chapter include entropic and other terms neglected at lower levels of theory. Water is modeled explicitly and molecular dynamics is used for sampling. Errors introduced by inadequate force fields and sampling techniques are difficult to isolate in complicated sites. The simple hydrophobic cavity site in lysozyme was a useful system to start testing the method. In our best effort, computed binding free energies on the set on known ligands had an RMS of 1.9 kcal/mol, and the known decoys were predicted as such. In blind prospective tests, the free energy methods correctly distinguished the known ligands and decoys. The RMS error of predicted binding free energies was only 0.6 kcal/mol. Structures for three of these ligands in complex with the cavity were determined, and they corresponded closely to the free energy predictions. In this work we realized the importance of including multiple ligand starting orientations, sampling side chain rotations, and an improved charge model in optimizing the free energy method. We also began to examine the effect of making

approximations such as holding the protein rigid as in docking. Not surprisingly, this made the method perform much worse.

Although this work appears as the third chapter in my thesis, it was actually my second paper to be published (*Journal of Molecular Biology* 2007). David Mobley and I contributed equally to this project. He did the computational work, while I performed the experimental work. I spent many weeks in the wet lab, learning how to do protein purification. While I had performed crystallography techniques prior to this project, it was a gratifying experience to work through the entire process on my own. Learning how to measure ligand binding with ITC was a major accomplishment for me; it also added great value to the results in the paper.

This work pushes the field a step forward towards predicting correct absolute binding free energies, and it highlights major challenges for more complex drug targets. The “top-down” rescoring approach allows us to pinpoint the most important terms for discriminating binders from non-binders, ranking the binders, and predicting their absolute binding free energies. Molecular docking can barely accomplish the first of these tasks, much less the latter two. The knowledge gleaned from this “top-down” rescoring approach is invaluable for directing the improvements of molecular docking and computational drug design tools in general.

### **3. Predicting Absolute Ligand Binding Free Energies to a Simple Model Site**

David L. Mobley,<sup>1,†</sup> Alan P. Graves,<sup>2,†</sup> John D. Chodera,<sup>2,‡</sup> Andrea C. McReynolds,<sup>1</sup>  
Brian K. Shoichet,<sup>1,\*</sup> and Ken A. Dill<sup>1,\*</sup>

<sup>1</sup>Department of Pharmaceutical Chemistry, and <sup>2</sup>Graduate Group in Biophysics,  
University of California San Francisco, 1700 4<sup>th</sup> Street, San Francisco, CA 94158-2330

<sup>†</sup> These authors contributed equally to this work

<sup>‡</sup> Current address: Department of Chemistry, Stanford University, Stanford, CA 94305-  
5080

\* Email addresses of corresponding authors:

BKS: shoichet@cgl.ucsf.edu, MPJ: matt.jacobson@ucsf.edu, DAC: case@scripps.edu

## **Abstract**

A central challenge in structure-based ligand design is the accurate prediction of binding free energies. Here, we apply alchemical free energy calculations in explicit solvent to predict ligand binding in a model cavity in T4 lysozyme. Even in this simple site, there are challenges. We made systematic improvements, beginning with single poses from docking, then including multiple poses, additional protein conformational changes, and using an improved charge model. Computed absolute binding free energies had an RMS error of 1.9 kcal/mol relative to previously determined experimental values. In blind prospective tests, the methods correctly discriminated between several true ligands and decoys in a set of putative binders identified by docking. In these prospective tests, the RMS error in predicted binding free energies relative to those subsequently determined experimentally was only 0.6 kcal/mol. X-ray crystal structures of the new ligands bound in the cavity corresponded closely to predictions from the free energy calculations, but sometimes differed from those predicted by docking. Finally, we examined the impact of holding the protein rigid, as in docking, with a view to learning how approximations made in docking affect accuracy and how they may be improved.

**Abbreviations:**

MM-GBSA, Molecular Mechanics-Generalized Born/Surface Area

MM-PBSA, Molecular Mechanics-Poisson-Boltzmann/Surface Area

L99A, Leu99 → Ala mutant of T4 lysozyme

CD, circular dichroism

PDB, the Protein Data Bank

ITC, isothermal titration calorimetry

**Keywords:**

Free energy calculation, docking, alchemical free energy, conformational change,  
isothermal titration calorimetry

## Introduction

A central problem in ligand discovery and design is the prediction of ligand--receptor binding free energies. Current methods cover a spectrum of physical rigor and computational cost. Among physics-based methods, physics-based docking and scoring is computationally the least expensive. In this approach, ligand orientations (poses) are assigned scores, related to the intermolecular interaction energy, and ranked relative to other poses and other ligands.<sup>133</sup> A few scoring functions include an explicit or implicit estimate of the desolvation free energy of the receptor and ligand.<sup>18</sup> Receptor flexibility, strain energies,<sup>133, 134</sup> and various entropies are usually neglected, as is any reference to the unbound protein and ligand states. These approximations put estimation of binding affinities well out of the reach of docking methods, although these methods can often correctly rank-order candidate molecules for testing.

At a higher level of theory are MM-GBSA/PBSA methods.<sup>135-137</sup> These methods estimate the absolute free energies of bound and unbound reference states. Enthalpies are estimated using average energies from a molecular mechanics force field, and combined with an entropy estimate and a solvation free energy from an implicit solvent model. The difficulty is that the binding free energy is a small difference between very large absolute free energies, requiring either very high accuracy in computing these large numbers, or cancellation of errors. Thus, while these approaches have had successes,<sup>136, 137</sup> they also have several drawbacks, such as sensitivity to details of the implicit solvent model used.<sup>138, 139</sup> and to the method used for estimating the entropy term. Such methods perform poorly on some test sets.<sup>140, 141</sup>

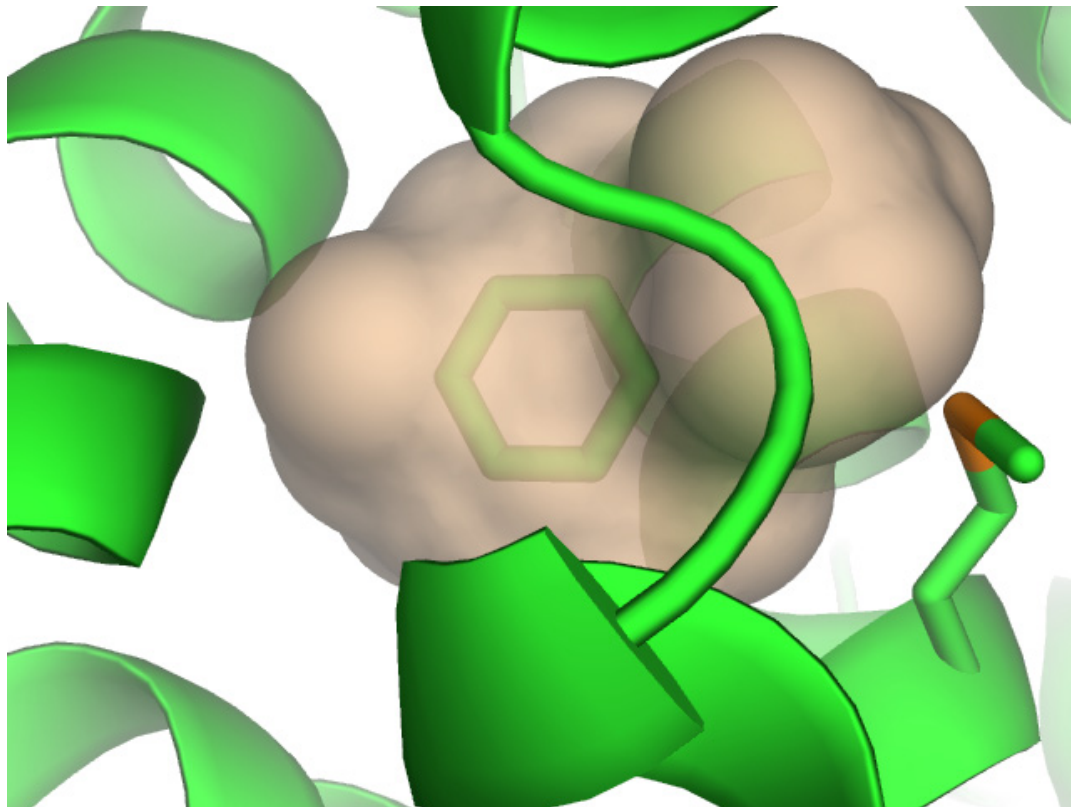
At the highest level of rigor are various free energy methods, including the alchemical free energy calculations described below, and PMF-based methods<sup>142, 143</sup> (for recent reviews of free energy methods, see works by Rodinger and Pomès, Kofke, and Shirts *et al.*).<sup>144-146</sup> Here, we focus on alchemical free energy methods, which evaluate

ratios of partition functions to estimate binding free energies, and thus include entropic and other contributions neglected at lower levels of theory. These methods, combined with some theoretical developments first laid out in the mid-1990s<sup>147-149</sup> and refined later,<sup>106</sup> now allow absolute binding free energies to be computed rigorously and exactly, provided that the molecular mechanics forcefield used accurately describes the underlying physics, and that enough sampling can be performed that the estimates of the relevant thermodynamic averages are converged.<sup>150</sup>

If, in principle, alchemical free energy calculations allow for exact prediction of binding free energies, the requirements of accurate force fields and adequate sampling introduce error into the computed free energies. This error is often difficult to isolate in the complicated environment of protein active sites. In such sites, failures of sampling or force fields are exacerbated by binding-induced conformational changes, titratable groups, metal ions, and ordered waters, among other complications. Furthermore, when sampling is inadequate, alchemical free energy methods can easily give biased results; for example, computed free energies are often sensitive to the choice of the initial receptor or ligand structure.<sup>105, 150-155</sup>

Here, to isolate sources of error, we study a highly simplified binding site using alchemical free energy methods and molecular dynamics. We focus on the binding of small aromatic ligands to the small, buried hydrophobic binding site in an engineered mutant of T4 lysozyme (the L99A site; Figure 1) that has been studied extensively experimentally,<sup>59, 63, 65, 104, 113, 156, 157</sup> with docking methods,<sup>59, 104</sup> and in some previous computational free energy studies.<sup>105, 106, 149, 158</sup> Here, we systematically evaluate the effect of various approximations on computed binding free energies. This model binding site provides a good starting point because it is simple and has been thoroughly characterized experimentally.





**Figure 1.** The model hydrophobic binding site in the L99A mutant of T4 lysozyme. The enclosed molecular surface of the cavity is shown (brown) as the crystallographic geometry of a bound benzene ligand (green), within the context of the overall structure of T4 lysozyme (green ribbons). The side-chain of Met102 is also shown for reference.

A second advantage of this model binding site is that it also provides an excellent opportunity for prospective predictions, since it is relatively easy to find new compounds that bind.<sup>59</sup> This is valuable, because it can be far easier to suggest explanations for previous observations than to actually make new predictions, and predictive ability provides a fundamental test for methods.

## Results

**Overview.** Here, we performed two sets of studies: Retrospective, in which we studied binding of ligands with previously measured affinities, and prospective, in which we predicted, in a blind test, the binding modes and affinities of several previously uncharacterized small molecules. After making predictions, we tested them experimentally, using isothermal titration calorimetry to measure affinities and X-ray crystallography to determine structures of the complexes.

**Retrospective studies: Comparison with previous experimental results.** We first computed binding free energies for a test set of 13 small neutral compounds using alchemical free energy calculations, as described in Methods (Table 1). Of these, binding affinities for 11 had previously been measured by ITC,<sup>63</sup> and two had previously been determined not to bind more strongly than an affinity of 10 mM using a thermal denaturation assay.<sup>104, 113</sup>

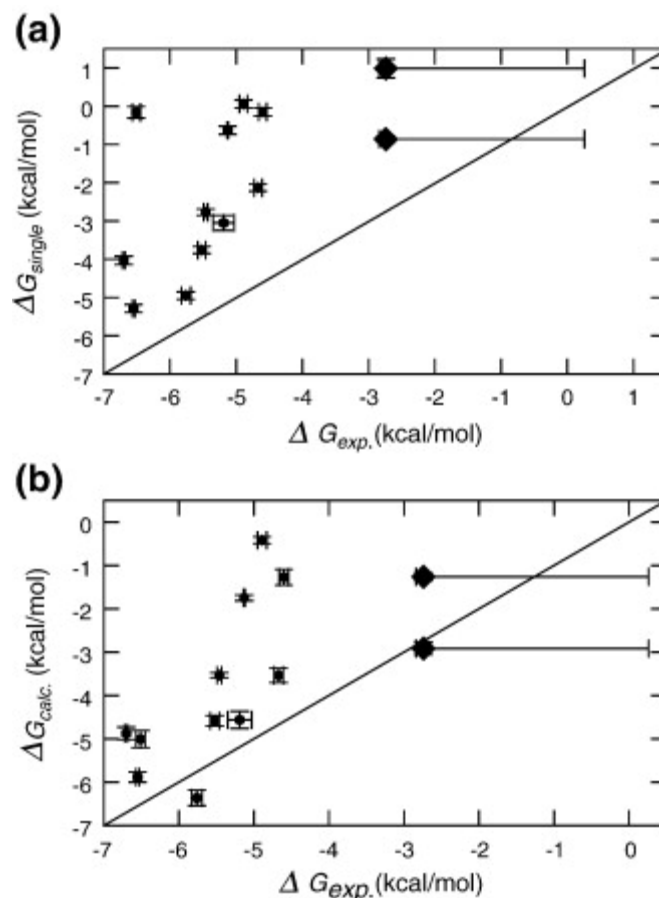
**Table 1.** Calculated and experimental binding free energies for ligands of the apolar binding site considered here.

Molecule	$\Delta G_{\text{exp}}^{\circ}$ (kcal/mol)	$\Delta G_{\text{single}}^{\circ}$ (kcal/mol)	$\Delta G_{\text{multiple}}^{\circ}$ (kcal/mol)	$\Delta G_{\text{calc}}^{\circ}$ (kcal/mol)	$\Delta G_{\text{calc}}^{\circ} - \Delta G_{\text{exp}}^{\circ}$ (kcal/mol)
2,3-Benzofuran	$-5.46 \pm 0.03$	$-2.77 \pm 0.08$	$-3.45 \pm 0.06$	$-3.53 \pm 0.06$	$1.93 \pm 0.07$
Benzene	$-5.19 \pm 0.16$	$-3.05 \pm 0.20$	$-4.53 \pm 0.20$	$-4.56 \pm 0.20$	$0.63 \pm 0.26$
Ethylbenzene	$-5.76 \pm 0.07$	$-4.95 \pm 0.10$	$-5.36 \pm 0.10$	$-6.36 \pm 0.18$	$-0.60 \pm 0.19$
Indene	$-5.13 \pm 0.01$	$-0.63 \pm 0.11$	$-1.56 \pm 0.06$	$-1.75 \pm 0.07$	$3.38 \pm 0.07$
Indole	$-4.89 \pm 0.06$	$0.06 \pm 0.10$	$-0.24 \pm 0.07$	$-0.42 \pm 0.08$	$4.47 \pm 0.10$
Isobutylbenzene	$-6.51 \pm 0.06$	$-0.16 \pm 0.15$	$-4.14 \pm 0.12$	$-5.01 \pm 0.20$	$1.50 \pm 0.21$
<i>n</i> -Butylbenzene	$-6.70 \pm 0.02$	$-4.03 \pm 0.11$	$-4.44 \pm 0.11$	$-4.87 \pm 0.14$	$1.83 \pm 0.14$
<i>n</i> -Propylbenzene	$-6.55 \pm 0.02$	$-5.29 \pm 0.10$	$-5.70 \pm 0.10$	$-5.88 \pm 0.11$	$0.67 \pm 0.12$
<i>o</i> -Xylene	$-4.60 \pm 0.06$	$-0.15 \pm 0.10$	$-0.56 \pm 0.10$	$-1.27 \pm 0.18$	$3.33 \pm 0.19$
<i>p</i> -Xylene	$-4.67 \pm 0.06$	$-2.13 \pm 0.09$	$-2.96 \pm 0.09$	$-3.54 \pm 0.17$	$1.13 \pm 0.18$
Toluene	$-5.52 \pm 0.06$	$-3.76 \pm 0.09$	$-4.17 \pm 0.09$	$-4.58 \pm 0.12$	$0.94 \pm 0.14$
Phenol	$> -2.74$	$-0.86 \pm 0.09$	$-1.27 \pm 0.09$	$-1.26 \pm 0.09$	N/A
2-Fluorobenzaldehyde	$> -2.74$	$0.99 \pm 0.25$	$-2.43 \pm 0.10$	$-2.92 \pm 0.14$	N/A
<i>Statistics</i>					
RMS error		$3.51 \pm 0.04$	$2.55 \pm 0.03$	$2.24 \pm 0.04$	
Correlation, <i>R</i>		$0.51 \pm 0.05$	$0.72 \pm 0.05$	$0.72 \pm 0.05$	

Experimental values (denoted by  $\Delta G_{\text{exp}}^{\circ}$ ) are according to Morton *et al.*,<sup>63</sup> except for 2-fluorobenzaldehyde and phenol, where no binding was observed (by a circular dichroism  $\Delta T_m$  upshift assay<sup>104, 113</sup>), giving only a lower bound on the binding free energy.

Calculated values shown are  $\Delta G_{\text{single}}^{\circ}$ , the free energy computed using only the single best-scoring docking orientation;  $\Delta G_{\text{multiple}}^{\circ}$ , the full computed binding free energy using all orientations considered; and  $\Delta G_{\text{calc}}^{\circ}$ , the computed binding free energy including multiple orientations and using the confine-and-release approach to account for protein conformational change at Val111. The final column is the difference from the experiment. At the bottom is the RMS error relative to the experiment across binders for each set of free energies, and the correlation coefficient, *R*, between calculated and experimental values. Experimental binding affinities were measured at 302 K; binding free energies were computed at 300 K. Calculated values used AM1-CM2 charges.

*Binding affinities are underestimated from single docking poses.* We started with a simple approach. We used the best-scoring docking pose for each compound as a starting structure from which to simply compute binding free energies using our standard free energy calculation protocol discussed in Methods. We previously found that this single-pose approach often results in ligands remaining trapped in the vicinity of their starting orientation on simulation timescales.<sup>150</sup> Thus, with this approach, the free energy calculations effectively become an expensive re-scoring of docking poses, including conformational averaging and entropic effects, but only for a single orientation. We present results using this approach as  $\Delta G_{\text{single}}^{\circ}$  (Table 1). In keeping with the approximations typically made in docking, we consider only *single* orientations for these results, meaning that we also neglect symmetry-equivalent orientations for molecules like toluene, phenol, and benzene, which in reality also contribute to binding.<sup>105, 106, 148, 150</sup> The RMS error for  $\Delta G_{\text{single}}^{\circ}$  relative to experiment is  $3.51 \pm 0.04$  kcal/mol, and the correlation coefficient ( $R$ ) between computed free energies and experiment is  $0.51 \pm 0.05$ . (These RMS and correlation calculations do not include the nonbinders, since free energies of association for these are not known.) (Figure 2(a)). This approach underestimates all the binding affinities. This is likely due to undersampling, a failure to adequately sample the most optimal binding conformations. Previous work had suggested this approach would fail in the case where the best docking pose is not the orientation that actually contributes most to binding,<sup>150</sup> so this outcome was expected.



**Figure 2.** Calculated binding free energies compared with the experiment. Calculated and experimental binding free energies are shown with error bars; the calculated error bars represent one standard deviation. The two points shown as larger diamonds are the nonligands phenol and 2-fluorobenzaldehyde; for these, only a lower limit on the experimental binding free energy is known, as denoted by a large experimental error bar to the right. The diagonal  $x = y$  line denotes perfect agreement with the experiment. (a) Calculated  $\Delta G_{\text{single}}^0$ , single-orientation binding free energies, including only the contribution from the single best docking orientation. (b)  $\Delta G_{\text{calc.}}^0$ , binding free energies, including all relevant ligand orientations and contributions from releasing Val111 from its kinetic confinement.

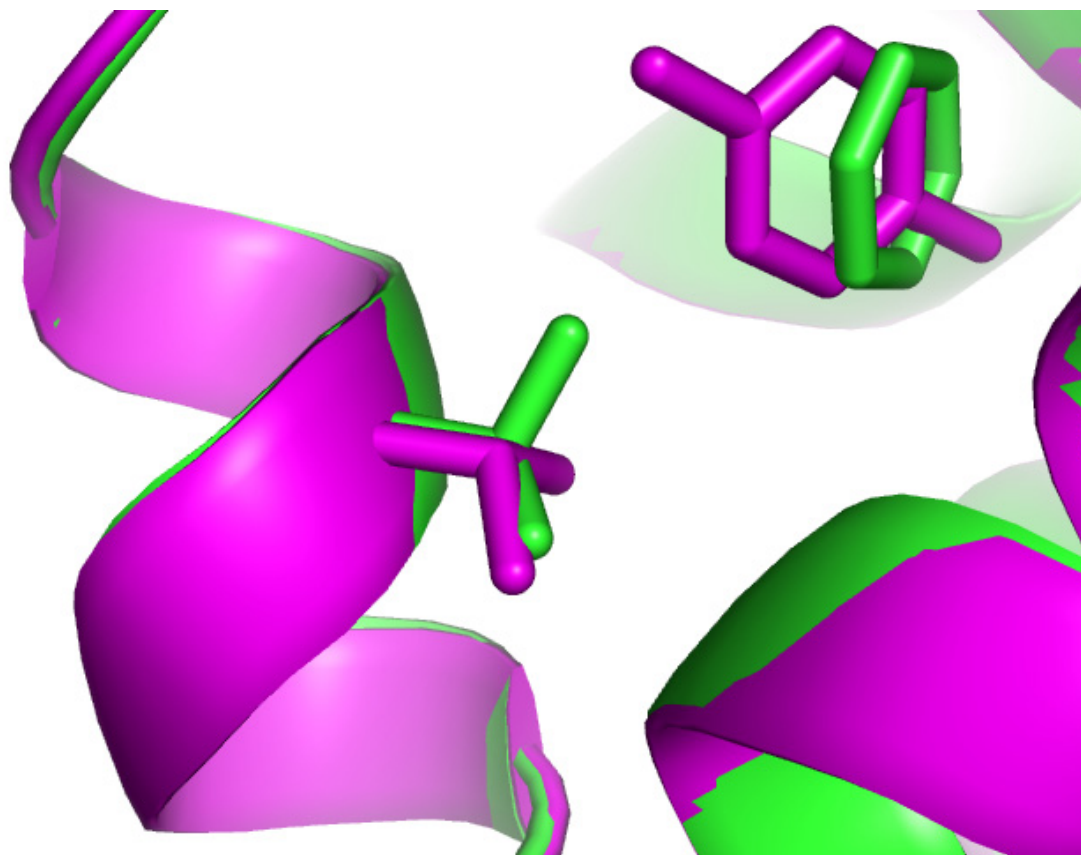
*Accounting for multiple potential bound orientations reduces the error in computed binding affinities.* Next, we account for the presence of multiple potential ligand binding modes separated by kinetic barriers. We compute binding free energies of different possible binding modes separately, and combine their contributions to get an overall binding free energy<sup>150</sup> (see also Simulation methods). We refer to these free energies, which also include the contributions of orientations related by symmetry, as  $\Delta G_{\text{multiple}}^{\circ}$  (Table 1). With this approach, the computed binding free energies are substantially closer to experiment in several cases (and are never worse) than those computed using the single-orientation approach, since occasionally the best-scoring pose is not the pose that contributes most to the binding free energy. This inclusion of these contributions reduces the RMS error in the computed free energies relative to experiment, from  $3.51 \pm 0.04$  kcal/mol to  $2.55 \pm 0.03$  kcal/mol, and raises the correlation coefficient,  $R$ , from  $0.51 \pm 0.05$  to  $0.72 \pm 0.05$ .

The improvements with this approach come for several reasons. For three of the ligands (indene, indole, and 2,3-benzofuran) multiple orientations are within  $kT$  of one another and all contribute substantially. For isobutylbenzene, the best pose from docking is not in the orientation that contributes most to binding, so including multiple candidate orientations results in inclusion of the dominant orientation. For the remainder of the compounds, improvements come from inclusion of symmetry number corrections. These issues have been addressed in more detail in work on a related binding site.<sup>150</sup> In general, it is extremely difficult to predict in advance whether multiple orientations may be relevant.

*Accounting for more protein conformational change further improves computed binding free energies.* The section above describes our treatment of relevant *ligand* orientations. However, the *protein* may also have relevant slow degrees of freedom which can be difficult to sample.<sup>113, 114</sup> Here, a key change is the reorientation of the Val111

sidechain observed in X-ray structures in response to ligands such as *n*-butylbenzene, isobutylbenzene, *o*-xylene, and *p*-xylene (Figure 3).<sup>113</sup> The energy barriers associated with this reorientation are large enough to prevent the sidechain from rotating on simulation timescales.<sup>114</sup> This leads to an apparent dependence of computed free energies on the initial protein structure used in simulations. For example, binding free energies that are computed from the *holo* protein structure are too negative (favorable) if the sidechain does not have time to re-orient as the ligand is removed, because the protein strain energy (the energetic cost of deforming the protein on binding) is not properly accounted for. On the other hand, if the *apo* protein structure is used, as we did here, binding free energies are too positive (unfavorable), as the ligand sterically clashes to some degree with the protein.<sup>114</sup> This dependence on the starting structure is simply due to kinetic trapping of the protein in conformations near its starting conformation.





**Figure 3.** Val111 reorients on ligand binding. Val111 is observed to adopt a different side-chain rotamer from the *apo* crystallographic structure in cocrystal structures with several different ligands. Shown here is the benzene-bound structure, green, which is virtually identical to the *apo* structure of the protein. Also shown is the *p*-xylene-bound structure in magenta. The sticks at the left show the reorientation of the Val111 side-chain on binding to *p*-xylene by roughly 120° relative to the benzene-bound and *apo* structures.

To overcome the kinetic trapping of Val111, we use a recently-developed “confine-and-release” framework to obtain correct binding free energies that are independent of the starting structure.<sup>114</sup> Specifically, when the Val111 remains trapped, computed binding free energies are really “confined” binding free energies, with Val111 confined to a particular orientation, so we use umbrella sampling to compute the free energy of releasing the valine from its confinement in the bound and unbound states. Here, this is accomplished by forcing sampling of alternative orientations using a harmonic biasing potential, and recovering the free energy landscape for this degree of freedom.<sup>114</sup> We do this for all of the compounds considered here, although for many compounds, only the *apo* orientation of the valine is found to be relevant, as observed experimentally.<sup>113</sup> This is a rigorous way to account for kinetic trapping. The confine-and-release framework is a generalization (to protein degrees of freedom) of the biasing potential approaches applied previously to ligands in a number of studies.<sup>105, 106, 142, 147, 149, 150, 155</sup>

The confine-and-release approach, which yields the total estimated binding free energy  $\Delta G_{\text{bind}}$ , further improves the agreement of computed binding free energies with experiment (Table 1 and Figure 2(b)). With this approach, the RMS error relative to experiment further decreases from  $2.55 \pm 0.03$  kcal/mol to  $2.24 \pm 0.04$  kcal/mol, while the correlation with experiment remains unchanged ( $R = 0.72 \pm 0.05$ ). There is significant improvement in the agreement with experiment for a number of the ligands, especially isobutylbenzene, *p*-xylene, *n*-butylbenzene, *o*-xylene, and ethylbenzene. As mentioned above, for the first four of these, Val111 re-orientation on binding is observed in the co-crystal structure.<sup>113</sup> For ethylbenzene, the deposited structure does not show the Val111 rotated relative to the benzene bound structure, but the electron density seems to allow the possibility of either orientation.<sup>113</sup> A prediction of this work is that, if the crystal structure with ethylbenzene can be solved at higher resolution, the Val111 sidechain

should be observed to adopt a conformation similar to that in the *p*-xylene bound structure.

The confine-and-release approach can be applied to a variety of protein conformational changes. Here, we chose to apply it specifically to a single Val111 dihedral. This choice was motivated by the fact that Val111 was previously observed (experimentally) to reorient on ligand binding.<sup>113</sup> Also, previous computational work led us to believe that sampling of this degree of freedom could be very slow.<sup>105</sup> We therefore examined our initial simulations to look for Val111 reorientation, and found the kinetic trapping described above.<sup>114</sup> This led us to apply the confine-and-release approach to this particular degree of freedom.

It is worth noting that this is not simply an issue of predicting a correct bound structure. Rather, using *either* the *apo* or *holo* structure leads to biased binding free energies (Tables 1 and 2) when the confine-and-release approach is not used. Only when we account for Val111 reorientation using confine-and-release do computed free energies become consistent between simulations beginning from *apo* and *holo* starting structures (see below).<sup>114</sup>

**Table 2.** Binding free energies calculated for selected ligands using their *holo* structures as a starting point.

Molecule	$\Delta G_{\text{multiple}}^{\text{o,holo}}$ (kcal/mol)	$\Delta G_{\text{calc}}^{\text{o,holo}}$ (kcal/mol)	$\Delta G_{\text{calc}}^{\text{o,holo}} - \Delta G_{\text{calc}}^{\text{o}}$ (kcal/mol)
Indene	$-1.44 \pm 0.07$	$-1.64 \pm 0.09$	$0.10 \pm 0.11$
<i>n</i> -Butylbenzene	$-9.17 \pm 0.13$	$-5.32 \pm 0.22$	$-0.45 \pm 0.26$
Isobutylbenzene	$-8.98 \pm 0.13$	$-4.80 \pm 0.21$	$0.20 \pm 0.29$
<i>p</i> -Xylene	$-7.27 \pm 0.09$	$-3.31 \pm 0.20$	$0.23 \pm 0.26$
<i>o</i> -Xylene	$-5.93 \pm 0.12$	$-1.91 \pm 0.21$	$-0.64 \pm 0.28$

Shown are the calculated binding free energies, beginning from the *holo* structures, with and without including any Val111 reorientation, and the difference from the calculated binding free energies using the *apo* structures as a starting point (including the Val111 reorientation) in Table 1.

*Binding free energies from holo structures agree with those from apo after accounting for Val111 reorientation.* After the confine-and-release calculations, there was still particularly poor agreement with experiment for several ligands, especially *o*-xylene, indene, indole, isobutylbenzene, and 2,3-benzofuran (the first three of these are the only binders in Figure 2(b) with computed binding free energies worse than -2 kcal/mol). One possible explanation is inadequate sampling, perhaps due to additional protein conformational rearrangements that are not being sampled. For example, for indene, isobutylbenzene, and *o*-xylene, helix F, which forms one side of the cavity, shifts around 2 Å on ligand binding, making the binding site larger.<sup>113</sup> Additionally, previous free energy calculations on the same system tended to overestimate binding free energies for some of these same compounds when beginning from the *holo* structures.<sup>105</sup>

However, inspection of simulation trajectories suggests that this helix motion is being sampled. As a more quantitative test, we repeated the calculations for selected compounds beginning from the *holo* structures. If computed binding free energies are different starting from *apo* and *holo* structures, even after applying the confine-and-release approach for Val111, it would indicate inadequate sampling. While computed binding free energies are significantly different for calculations started from the *apo* and *holo* structure *before* accounting for Val111 reorientation, the differences are essentially negligible (within uncertainty) when the confine-and-release approach is used to account for this change (Table 2). (The largest difference, using the confine-and-release approach, is for *o*-xylene,  $-0.64 \pm 0.28$  kcal/mol; since the uncertainty represents one standard deviation, this is still only a  $2\sigma$  variation). This implies that sampling of these conformational changes is probably sufficient and that the error lies elsewhere.

Free energies computed using the *holo* starting structures also show that the *holo* protein structure of several of these ligands is unfavorable by roughly 4 kcal/mol in the absence of bound ligand (Table 2).<sup>114</sup> This is presumably because of steric clashes with

the protein, and is the reason why only some of the ligands induce this conformational change on binding.

*The AM1-BCC charge model further increases the accuracy of binding free energies.* Next, we considered another possible source of error: the simulation parameters. There are different methods for assigning partial charges for small molecules.<sup>159</sup> In the work reported above, we used AM1-CM2<sup>72</sup> partial atomic charges for the small molecules, as in docking studies. However, we found previously that AM1-BCC charges performed better than AM1-CM2 charges for hydration free energies, perhaps because they are more similar to the HF 6/31G\* charges the force field was parameterized with.<sup>159</sup> Therefore, we tested the AM1-BCC charges here as well. Table 3 shows that AM1-BCC charges further reduce the RMS error between computed and experimental binding free energies from  $2.24 \pm 0.04$  to  $1.89 \pm 0.04$  kcal/mol, and the correlation coefficient,  $R$ , increases from  $0.72 \pm 0.05$  to  $0.79 \pm 0.07$ .

**Table 3.** Calculated and experimental binding free energies for ligands of the apolar binding site considered here, as in Table 1 except using AM1-BCC charges.

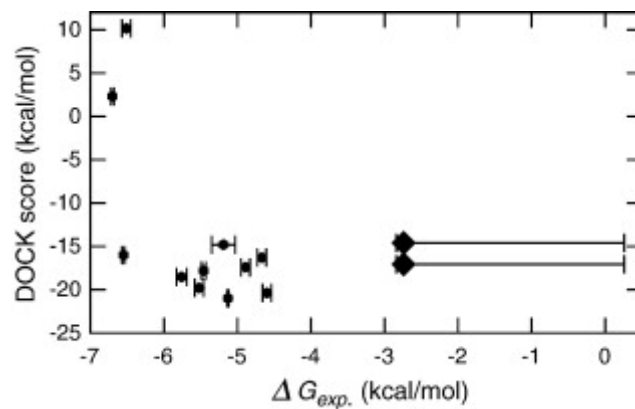
Molecule	$\Delta G_{\text{exp}}^{\circ}$ (kcal/mol)	$\Delta G_{\text{BCC}}^{\circ}$ (kcal/mol)	$\Delta G_{\text{BCC}}^{\circ} - \Delta G_{\text{exp}}^{\circ}$ (kcal/mol)	$\Delta G_{\text{AM1-CM2}}^{\circ} - \Delta G_{\text{BCC}}^{\circ}$ (kcal/mol)
2,3-Benzofuran	$-5.46 \pm 0.03$	$-3.66 \pm 0.06$	$1.80 \pm 0.06$	$0.13 \pm 0.08$
Benzene	$-5.19 \pm 0.16$	$-3.95 \pm 0.20$	$1.24 \pm 0.26$	$-0.61 \pm 0.28$
Ethylbenzene	$-5.76 \pm 0.07$	$-5.82 \pm 0.14$	$-0.06 \pm 0.16$	$-0.54 \pm 0.23$
Indene	$-5.13 \pm 0.01$	$-1.63 \pm 0.07$	$3.50 \pm 0.07$	$-0.12 \pm 0.09$
Indole	$-4.89 \pm 0.06$	$-1.37 \pm 0.10$	$3.52 \pm 0.12$	$0.96 \pm 0.13$
Isobutylbenzene	$-6.51 \pm 0.06$	$-8.09 \pm 0.18$	$-1.58 \pm 0.19$	$3.09 \pm 0.27$
<i>n</i> -Butylbenzene	$-6.70 \pm 0.02$	$-5.70 \pm 0.20$	$1.00 \pm 0.20$	$0.83 \pm 0.25$
<i>n</i> -Propylbenzene	$-6.55 \pm 0.02$	$-5.44 \pm 0.10$	$1.11 \pm 0.11$	$-0.44 \pm 0.16$
<i>o</i> -Xylene	$-4.60 \pm 0.06$	$-3.23 \pm 0.25$	$1.37 \pm 0.25$	$1.96 \pm 0.31$
<i>p</i> -Xylene	$-4.67 \pm 0.06$	$-3.59 \pm 0.12$	$1.08 \pm 0.14$	$0.05 \pm 0.21$
Toluene	$-5.52 \pm 0.06$	$-4.07 \pm 0.10$	$1.45 \pm 0.12$	$-0.51 \pm 0.16$
Phenol	$> -2.74$	$-1.07 \pm 0.20$	N/A	$-0.19 \pm 0.22$
2-Fluorobenzaldehyde	$> -2.74$	$-3.14 \pm 0.13$	N/A	$0.22 \pm 0.19$
<i>Statistics</i>				
RMS error		$1.89 \pm 0.04$		
Correlation, <i>R</i>		$0.79 \pm 0.07$		

Shown are  $\Delta G_{\text{BCC}}^{\circ}$ -full binding free energies done with AM1-BCC including contributions from multiple ligand orientations and any Val111 reorientation. These are equivalent to  $\Delta G_{\text{calc}}^{\circ}$  from Table 1 but done with AM1-BCC charges. Also shown are the differences between AM1-BCC results and the experiment (next to last column), and between AM1-BCC and AM1-CM2 results (last column). When the values in the last two columns have the same sign, AM1-BCC charges improved the agreement with the experiment. At the bottom is RMS error relative to the experiment across binders for each set of free energies, and the correlation coefficient, *R*, between calculated and experimental values.  
using AM1-BCC charges

*Alchemical methods are more accurate than docking.* One major challenge for docking methods is to discriminate between binders and non-binders. We have included two known non-binders (with affinities worse than 10 mM) in the set of molecules examined here: phenol and 2-fluorobenzaldehyde. For these two compounds, computed binding free energies indicate only weak affinity:  $-2.9 \pm 0.1$  kcal/mol and weaker (more positive) (Table 1). A 10mM detection threshold in binding affinity corresponds to a binding free energy of roughly  $-2.7$  kcal/mol. Thus, the computed binding free energies for these two compounds are at the detection limit, essentially consistent with the experimental observation that they are non-binders.

Our free energy calculations are computationally expensive. Are the results any more accurate than those that can be obtained from molecular docking? As shown in Figure 4, DOCK scores for the ligands studied here correlate poorly with experimental binding free energies. In fact, they are anti-correlated ( $R = -0.69$ ), the opposite of what one would like. Moreover, the two non-binders have DOCK scores similar to those for the majority of the true ligands (and much more favorable scores than several ligands), hence it is impossible to discriminate between binders and non-binders. In fairness to docking, it is worth noting that these nonbinders were included in our test set *because* they have proven challenging for docking to discriminate from the binders. Also, the first goal of docking is to separate likely from unlikely ligands, and it does seem to be performing remarkably well in this binding site, where about 80% of the top 100 docking hits would probably bind. Additionally, we find that docking also performs quite well in this site at generating sterically reasonable potential bound orientations. That said, the free energy calculations give substantially better affinity estimates and correlations than docking does, and are better at recognizing nonbinders.



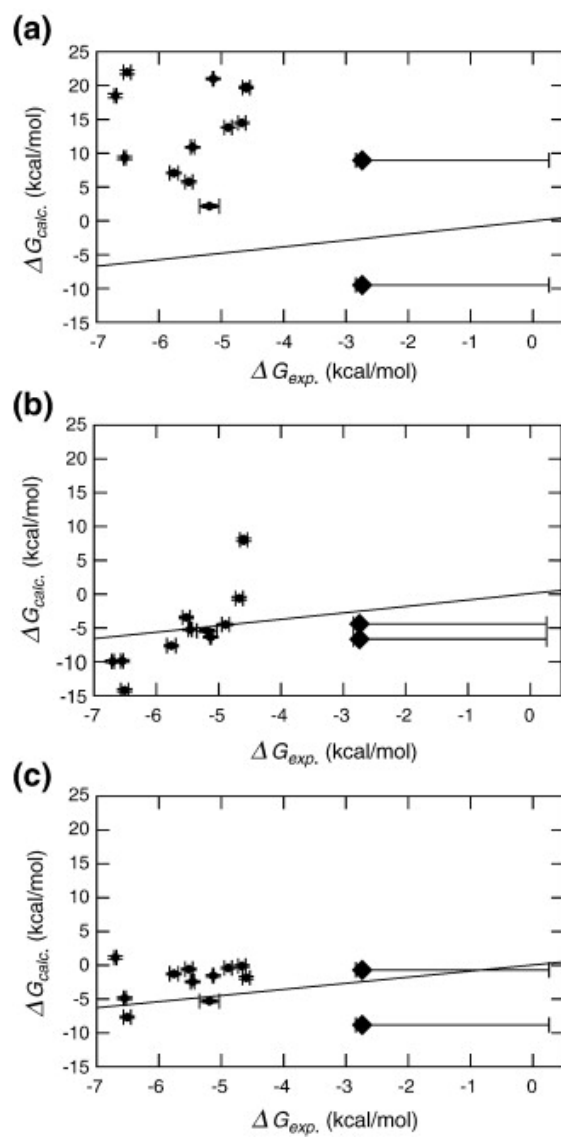


**Figure 4.** DOCK scores for the best-ranked pose for each molecule versus experimental binding free energies. The correlation coefficient ( $R$ ) is  $-0.69$ , meaning that compounds that DOCK predicts should bind strongly tend to bind weakly. Additionally, the two nonbinders have similar DOCK score to a number of the binders.

The docking results discussed here used the benzene-bound protein structure, which is virtually identical to the *apo* structure. However, docking to alternate protein conformations seems not to result in significant improvements in quality of docking results, except when *many* different crystallographic protein conformations are used.<sup>65</sup>

**A large source of error in docking is the rigid-protein approximation.**

Docking typically treats proteins as rigid. How big is the error introduced by this assumption? To test this, we held the protein rigid and repeated our free energy calculations, including the effects of ligand symmetries and multiple ligand orientations (and using AM1-CM2 charges). This led to essentially zero correlation ( $R = -0.05 \pm 0.09$ ) between computed free energies and experimental values, and an RMS error of  $19.78 \pm 0.06$  kcal/mol (Figure 5(a)).



**Figure 5.** Comparison of calculated and experimental binding free energies with the protein held rigid. (a) Binding free energies with the protein completely rigid. The RMS error relative to the experiment is  $19.78 \pm 0.06$  kcal/mol and the correlation coefficient ( $R$ ) is  $-0.05 \pm 0.09$ . (b) Binding free energies with the whole protein minimized separately for each ligand. The RMS error relative to the experiment is  $4.92 \pm 0.07$  kcal/mol and the correlation coefficient ( $R$ ) is  $0.82 \pm 0.09$ . (c) Binding free energies with only the binding site minimized for each ligand. The RMS error relative to the experiment is  $4.06 \pm 0.06$  kcal/mol and the correlation coefficient ( $R$ ) is  $0.32 \pm 0.08$ . The  $x = y$  indicates perfect agreement with the experiment.

As a simple improvement on this rigid protein approximation, we also allowed the protein to relax to a different structure for each ligand. First, we minimized the entire protein in the presence of each ligand, using the initial docking geometry, and subsequently held the protein rigid during the simulations. This resulted in a correlation ( $R$ ) of  $0.82 \pm 0.09$  and an RMS error of  $4.92 \pm 0.07$  kcal/mol relative to experiment (Figure 5(b)). Second, we minimized only a region of the protein around the binding site (Binding free energies to a rigid protein) in the presence of each ligand, before holding the protein fixed during the simulations. This resulted in a correlation ( $R$ ) of  $0.32 \pm 0.08$  and an RMS error of  $4.06 \pm 0.06$  kcal/mol relative to experiment (Figure 5(c)).

Overall, it appears that keeping the protein rigid while estimating binding free energies is detrimental to binding free energy estimation, even if minimization is performed separately for each ligand.

### **Prospective studies: Predictions and experimental tests.**

*Distinguishing binders from nonbinders.* We also performed a blind test of these free energy methods. We selected five small molecules that were among the top-scoring molecules from a docking screen of a compound library (using protocols described previously).<sup>104</sup> We calculated binding free energies in the same manner as above, and compared the resulting dissociation constants with the detection threshold of 10 mM for the experimental thermal denaturation assay. We predicted that four of the molecules would bind and one would not<sup>‡</sup> (the 10 mM threshold fell just between affinities for two of the compounds). We then tested these predictions experimentally using the upshift in thermal denaturation, in which the melting temperatures of the protein in the presence and absence of the ligand were compared.<sup>156</sup> All molecules were tested in their neutral forms, using either circular dichroism or fluorescence to monitor the transition from the

---

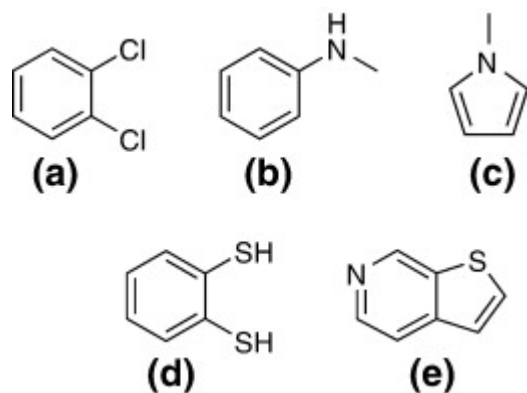
<sup>‡</sup> Initial predictions were made with AM1-CM2 charges, but AM1-BCC charges were tried later and led to the same outcome.

folded to the unfolded state, and resulting  $T_m$  values were compared to that of the *apo* protein. All melts occurred reversibly in a manner consistent with two-state unfolding. 1,2-dichlorobenzene, *n*-methylaniline, 1-methylpyrrole and 1,2-benzenedithiol increased the  $T_m$  significantly, between 1.0 and 2.9 °C (Table 4). Conversely, thieno-[2,3c]pyridine was not observed to increase the  $T_m$ , even at 2.5 mM concentration, consistent with the predictions of the free energy calculations. In contrast, docking had predicted all five would bind, but the free energy methods correctly identified the nonbinder (thieno[2,3-c]pyridine) (Table 4; Figure 6).

**Table 4.** Novel compounds for which predictions were made and later tested experimentally.

Molecule	DOCK score (kcal/mol)	Alchemical predicition <sup>a</sup>	$\Delta T_m$ (°C)	Experiment	$\Delta G_{\text{calc}}^{\circ}$ <sup>b</sup> (kcal/mol)	$\Delta G_{\text{exp}}^{\circ}$ <sup>c</sup> (kcal/mol)
1,2-Dichlorobenzene	-19.99	Binder	2.90	Binder	$-5.66 \pm 0.15$	-6.37
<i>n</i> -Methylaniline	-17.29	Binder	1.00	Binder	$-5.37 \pm 0.11$	-4.70
1-Methylpyrrole	-15.27	Binder	2.20	Binder	$-4.32 \pm 0.08$	-4.44
1,2-Benzenedithiol	-18.51	Binder	2.50	Binder	$-2.79 \pm 0.13$	N.D.
Thieno[2,3- <i>c</i> ]pyridine	-18.81	Nonbinder	-0.40	Nonbinder	$-2.56 \pm 0.07$	N.D.

DOCK scores suggested that all five should bind. Alchemical free energy calculations were initially used to predict whether or not these molecules would bind, then  $\Delta T_m$  values were determined experimentally to test these predictions. Following this, final binding free energy predictions ( $\Delta G_{\text{calc}}^{\circ}$ ) were tested experimentally with ITC; results are as shown ( $\Delta G_{\text{exp}}^{\circ}$ ). The RMS difference between predicted  $\Delta G^{\circ}$  and experiment for the three compounds tested with ITC is 0.57 kcal/mol. N.D. indicates that the measurements were not done. <sup>a</sup> Initial predictions were made using AM1-CM2 charges, but the outcome was unchanged with AM1-BCC charges. <sup>b</sup> Before doing ITC, predictions were refined using AM1-BCC charges, which testing had indicated gave higher accuracy. <sup>c</sup> Calorimetry was done at 283 K.



**Figure 6.** Five compounds for which binding predictions were made. (a) 1,2-dichlorobenzene; (b) *n*-methylaniline; (c) 1-methylpyrrole; (d) 1,2-benzenedithiol; and (e) thieno[2,3-*c*]pyridine.

*Predicting bound orientations was successful.* We then obtained crystal structures (Table 5) to determine how well these free energy calculations could predict the bound ligand conformations. We soaked three of the ligands into L99A lysozyme protein crystals. The crystals diffracted to between 1.7 Å and 2.07 Å on a home source. In all three structures, initial  $F_o-F_c$  electron density unambiguously identified the orientation of the ligand in the site; for dichlorobenzene, two orientations of the ligand were apparent.

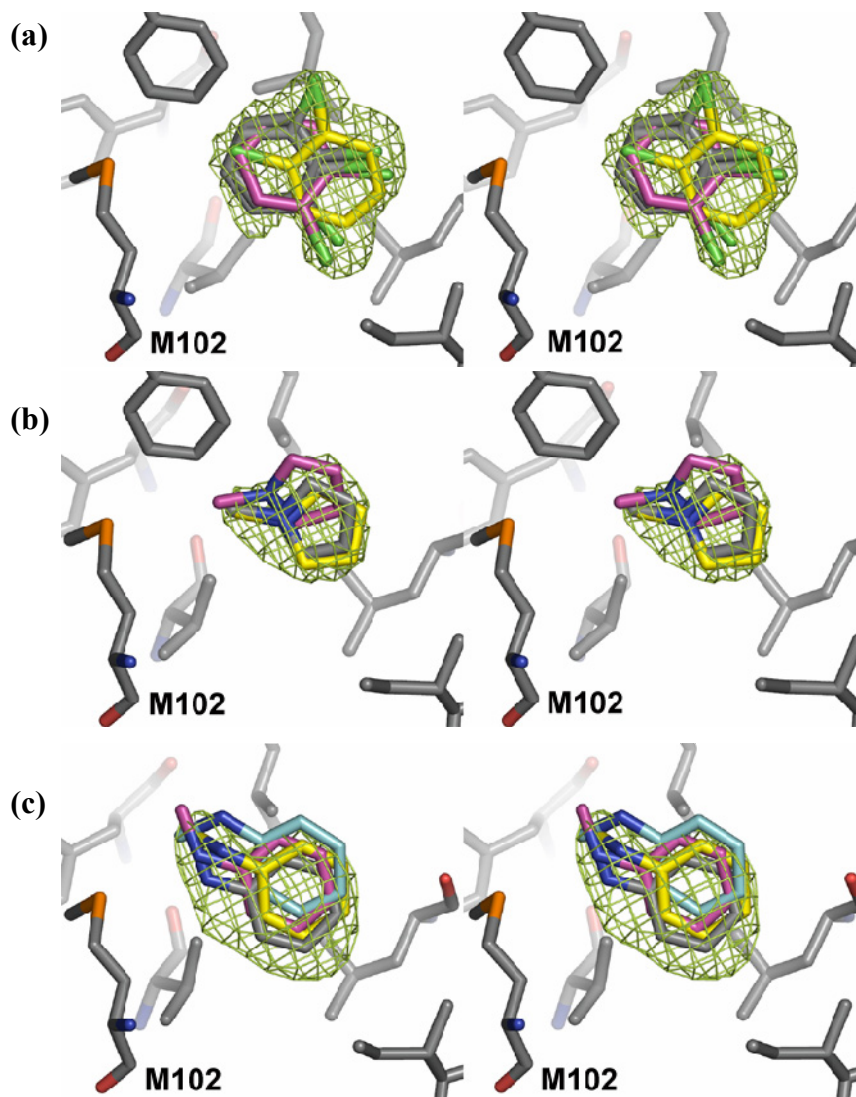


**Table 5.** X-ray data collection and refinement.

	Ligands bound to L99A		
	1,2-dichlorobenzene	<i>n</i> -methylaniline	1-methylpyrrole
Cell dimensions			
$a=b$ (Å)	60.2	59.9	60.2
$c$ (Å)	97.0	96.1	96.4
Resolution (Å)	1.70 (1.76) <sup>a</sup>	2.07 (2.14) <sup>a</sup>	1.94 (2.01) <sup>a</sup>
Reflections	18469 (2263) <sup>a</sup>	12102 (1169) <sup>a</sup>	15355 (1484) <sup>a</sup>
R <sub>merge</sub> (%)	9.8 (64.5) <sup>a</sup>	13.5 (52.8) <sup>a</sup>	11.2 (56.9) <sup>a</sup>
Completeness (%)	99.8 (99.9) <sup>a</sup>	95.0 (92.8) <sup>a</sup>	99.4 (98.4) <sup>a</sup>
$\langle I \rangle / \langle \sigma(I) \rangle$	9.8 (2.3) <sup>a</sup>	7.4 (2.3) <sup>a</sup>	8.7 (2.1) <sup>a</sup>
R-factor (%)	19.8	21.9	19.1
Resolution range (Å)	50.0-1.70	50.0-2.07	50.0-1.94
$\Delta_{\text{bond}}$ lengths (Å)	0.008	0.007	0.009
$\Delta_{\text{bond}}$ angles (°)	1.004	0.916	1.074
PDB code	2OTY	2OTZ	2OU0

<sup>a</sup> Values in parentheses are for the highest resolution shell.

In parallel, we predicted dominant bound orientations for each of these ligands (Simulation methods). Then we determined the structures for these three ligands, and compared with our predicted structures, the best predicted DOCK poses, and the electron density (Figure 7).



**Figure 7.** Predicted and experimental ligand orientations. Stereo images comparing the experimental and predicted poses for three ligands bound to L99A. (a) The two observed configurations of 1,2-dichlorobenzene, structure determined to 1.70 Å resolution. (b) 1-Methylpyrrole, structure determined to 1.94 Å resolution and (c) *n*-methylaniline, structure determined to 2.07 Å resolution. The crystallographic carbon atoms of protein residue M102 and each ligand are colored grey. The carbon atoms of the docking predictions are colored yellow, and the carbon atoms of the free energy predictions are colored magenta. The carbon atoms of the second free energy prediction in (c) are colored cyan. The  $F_o - F_c$  density maps are contoured at  $3\sigma$  (green mesh).

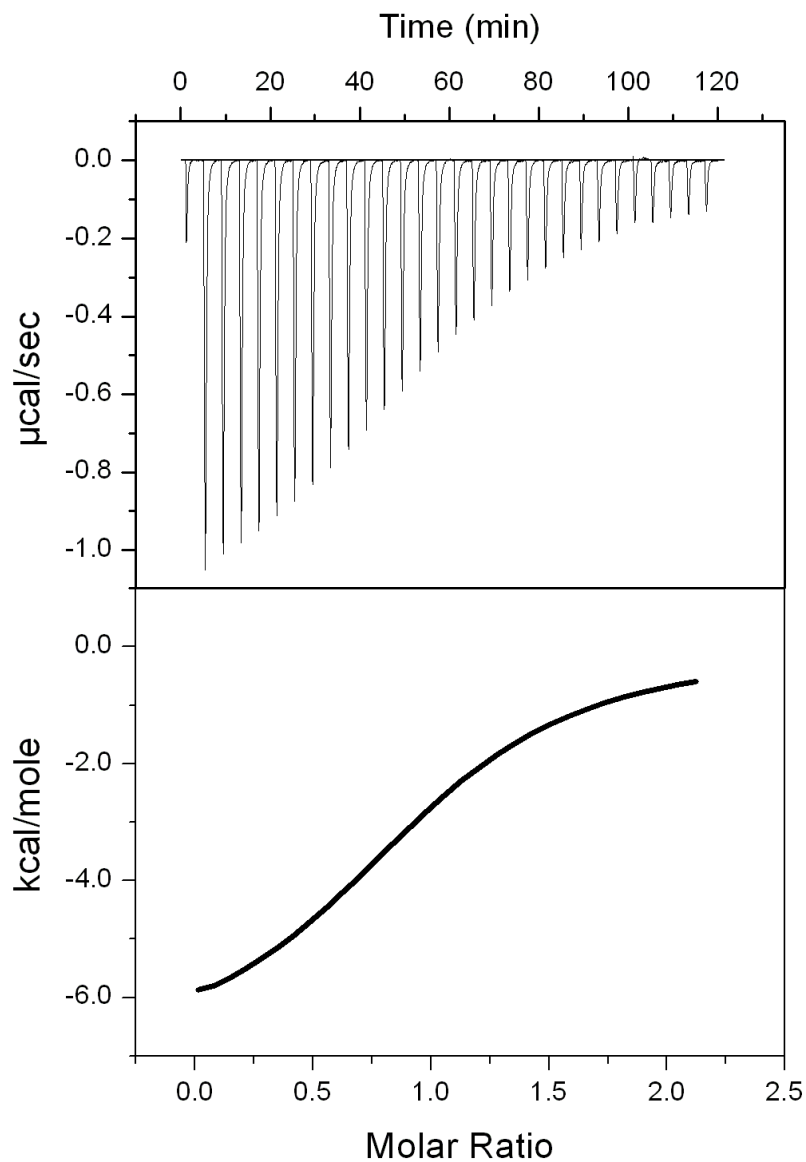
For 1-methylpyrrole, the X-ray, docking, and molecular dynamics poses were quite similar (Figure 7). The particular molecular dynamics snapshot that was selected as representative appears slightly twisted relative to the other two structures, but this is simply due to the arbitrary selection of a single MD conformation from an ensemble. The X-ray pose falls well within the range of structures sampled by the simulation from which this snapshot was chosen (and the underlying electron density is consistent with a range of structures seen in simulation). The RMSD between the docking pose and X-ray structure is 0.39 Å and that between the free energy snapshot and X-ray is 0.94 Å.

For *n*-methylaniline, free energy methods predicted two orientations, with essentially equal occupancy probabilities (Figure 7). The X-ray and docking poses match well with one of these orientations, but there is no evidence in the electron density for the second orientation. The two orientations differ only by rotation around the C1–N bond. The RMSD between the docking pose and X-ray is 0.63 Å, and that between the lower RMSD free energy snapshot and X-ray is 0.69 Å (the higher RMSD orientation is 1.29 Å away from X-ray).

For 1,2-dichlorobenzene, two separate orientations were observed in the crystal structure, differing by a rotation of around 60° in the plane of the aromatic ring. DOCK failed to properly identify *either* of these orientations as the best-scoring pose, whereas our free energy methods picked out one but indicated that the second was energetically unfavorable. In particular, the ligand would occasionally transition into this alternate orientation, but it would typically remain only transiently. We further tested this by conducting a separate set of calculations where the ligand was restrained to remain in the alternate orientation, but, with our parameter set, it appears substantially less favorable for binding than the orientation predicted to dominate using our free energy methods. This could be a force-field failure, inadequate sampling, or a difference between experimental and simulation conditions. The docking pose is 2.8-2.9 Å away from the X-

ray orientations, while the free energy snapshot is only 0.77 Å away from the most similar X-ray orientation.

*Predicting binding affinities.* Next, we predicted binding free energies and then measured them by isothermal titration calorimetry. Since the AM1-BCC charge model worked best in retrospective studies, we used these charges for binding free energy predictions. Ligand titrations gave easily modeled curves using a low  $c$ -value protocol (Figure 8).<sup>160</sup> Not only did binding free energy calculations give the correct rank ordering of binders, but computed binding free energies for these compounds were remarkably accurate (an RMS error of  $0.57 \pm 0.09$  kcal/mol, Table 4).



**Figure 8.** Representative ITC data. Data and fit for T4 lysozyme L99A (0.063 mM) titrated with 1,2-dichlorobenzene ( $\sim 0.6$  mM). An initial injection of  $2.5 \mu\text{l}$  was followed by 29 injections of  $10 \mu\text{l}$  of the ligand solution made every 2.5 min into the 1.4 ml reaction cell. After subtraction of blank runs, titrations were fit as described under Experimental methods to obtain the results in Table 4.

## Discussion

### Accuracy of free energy calculations in retrospective and prospective studies.

Alchemical free energy calculations using molecular dynamics can be used to compute fairly accurate binding free energies of ligands in the T4 lysozyme L99A binding site, with an RMS error of  $1.89 \pm 0.04$  kcal/mol in retrospective tests. This is a much higher accuracy than docking, where scores were inversely correlated with experiment, at least among the top-scoring ligands here. Admittedly, the docking program, DOCK, was never designed to predict binding affinities, and performs remarkably well at ranking likely ligands highly in large libraries.<sup>104</sup> Also, in these calculations, we are comparing with previously known results. A more rigorous test is to compare genuinely new predictions on untested candidate ligands with subsequent experiment.

Therefore, in a blind test, we predicted affinities and binding orientations for five previously uncharacterized compounds predicted by DOCK to bind, then tested these predictions experimentally. With alchemical free energy calculations, we correctly recognized the one non-binder, accurately predicted ligand bound orientations, and quantitatively predicted binding free energies. In each of these areas, free energy methods agreed better with experiment than docking did. Free energy methods, unlike docking, also correctly ranked the ligand binding affinities in these prospective tests. Thus, it appears that alchemical free energy methods can be truly predictive and can rescue docking failures.

The approach described here, including the retrospective study of previously published data, required no knowledge of the bound structure of the protein and ligand, and used the *apo* protein structure. Previous work on this same binding site has required an accurate bound structure of the complex of interest as a starting point.<sup>105</sup>

**Essential ingredients for accuracy.** This present study is limited to a simplified model binding site, where many complications of other binding sites are absent. The L99A cavity studied here has no interface with bulk water, no ordered waters to displace, is small, and the dominant interactions appear to be mainly non-polar. Nevertheless, the use of this system has allowed us to be systematic in isolating and solving various sampling problems. We identified several keys to obtaining accurate binding free energies: First, multiple potential ligand orientations must be included; one cannot rely on the single top docking pose to be the dominant ligand orientation. There can be large energetic barriers between different ligand orientations which make timescales for orientational change long compared to simulation times. Second, even seemingly small protein conformational changes on ligand binding, such as the reorientation of a single sidechain, Val111, can be difficult to sample correctly in free energy calculations, yet it is essential to include these conformational changes to get correct binding free energies.

On this second point, it is interesting to note that previous computational work on this same binding site gave free energies that were more than 2 kcal/mol too negative for four of 10 known binders in calculations initiated from the *holo* structures.<sup>105</sup> We performed similar free energy calculations for these compounds, found that our calculations overestimated binding affinities for the same ligands, despite the fact that we use a different force field and different parameters, but only when we failed to account for the free energy associated with Val111 reorientation. Indeed, in the previous work, results were sensitive to the starting orientation of the Val111 sidechain for at least one ligand, indicating that this sidechain degree of freedom was inadequately sampled.<sup>105</sup> Our results indicate that if the free energy associated with this reorientation is not included, it can bias computed binding free energies by as much as 4 kcal/mol.

Given that these small conformational changes can contribute so substantially to overall binding free energies, it will likely prove essential to develop improved methods



for computing free energies associated with larger-scale protein conformational changes, such as loop motions on ligand binding.

**Lessons for docking.** To address the rigid protein assumption typically used in docking, we also tried free energy calculations with the protein held rigid. Using a rigid *apo* structure for all ligands resulted in very large errors (RMS error 20 kcal/mol; zero correlation). Minimizing the protein separately in the presence of each ligand worked better, but RMS errors remained high (above 4 kcal/mol), and the approach lacked the ability to recognize non-binders. Apparently, for this binding site, holding the protein rigid cannot easily produce binding affinities that agree quantitatively (even within 4 kcal/mol RMS error) with experiment, but strategies involving minimization of the protein can provide some improvement over treating it as completely rigid. But for high accuracy, it may be necessary to include not only protein conformational change, but a correct accounting for the free energy costs associated with these protein conformational changes -- which can be substantial, even at the single sidechain level. Lastly, our results indicated that higher-quality charges can lead to substantial improvements in binding affinities; thus, the AM1-BCC charges that performed best here may also be a better choice for docking.

**Conclusions.** Overall, our results indicate that free energy methods are reaching the point where they can be useful when used predictively. However, in the relatively simple system examined here, this reasonably high level of accuracy depends on carefully accounting for the presence of multiple potential ligand bound orientations, and the possibility of protein conformational changes on ligand binding. In principle extremely long molecular dynamics simulations can handle both of these, but in practice, the computational cost of such simulations is often prohibitive. For now, treating both problems requires deliberate sampling of the relevant degrees of freedom, and so some

pre-knowledge of these degrees of freedom. We suspect that challenges observed in this model binding site will be found in biologically relevant binding sites as well. Despite these limitations, alchemical free energy methods hold great promise, both in predictive power, and in guiding improvement of more approximate physics-based methods.

## Materials and Methods

### Simulation Methods.

*Overview.* We begin with the benzene-bound crystal structure of the T4 lysozyme L99A mutant (which is virtually identical to the *apo* structure, 1L90, from which it has an RMSD of 0.14 Å), dock our ligands into the binding site, and determine which poses or orientations are kinetically stable and distinct. We then begin free energy calculations from stable, distinct orientations, as described previously,<sup>150</sup> to compute binding free energies. Finally, for each ligand, we combine contributions from these different orientations to rigorously estimate binding free energies.

*System preparation.* Except where indicated otherwise, the initial protein structure for molecular dynamics simulations and free energy calculations is the benzene-bound structure of the L99A lysozyme mutant, which is essentially identical to the *apo* structure. This was prepared with the GROMACS<sup>161, 162</sup> 3.3 utility PDB2GMX with default protonation states, using a GROMACS port<sup>163</sup> of the AMBER 96<sup>164</sup> force field. Since the cavity that makes up the binding site is completely hydrophobic without any nearby titratable groups, these protonation states present no difficulties. Following preparation, the protein was placed in a dodecahedral simulation box and surrounded by roughly 6,000 water molecules which were pre-equilibrated for 1 ns with the protein held fixed prior to the equilibration of the full system, which is discussed below.

*Docking.* We used DOCK3.5.54 to fit the molecules of interest into the protein structure (Tables 1 and 4). We retained all of the generated poses (numbering in the thousands) and scores of the molecules, then sorted these by score. We then began with the best scoring pose and worked towards the worst, retaining every pose that was different by more than 2 Å RMSD from a better scoring pose, to generate a set of the best

scoring, distinct ligand orientations. This typically resulted in 10–40 distinct poses, of which we retained only the group of top-scoring poses (typically 3–8).

*Identifying candidate orientations.* From these poses, we generated general AMBER force field<sup>122</sup> parameters for each ligand using ANTECHAMBER version 1.2.4,<sup>121</sup> and AM1-CM2 charges<sup>71</sup> as discussed previously.<sup>150</sup> These charges were employed in docking studies on the same system,<sup>104</sup> and we sought to separate parameter differences from methodology differences as much as possible. We also present results in this work where we use AM1-BCC<sup>124, 165</sup> charges computed with ANTECHAMBER.

From the resulting small-molecule AMBER topology and coordinate files, we generated GROMACS topology and coordinate files using the amb2gmx conversion utility<sup>§</sup>. These ligand topologies and coordinates were then merged with those for the pre-solvated protein system prior to simulation.

To further reduce the number of ligand orientations we consider in free energy calculations, we initiated separate 1 ns molecular dynamics simulations from all candidate orientations to identify those that are kinetically distinct.<sup>150</sup> We only retained one orientation of each set of orientations that interconvert easily within simulation timescales. This typically resulted in 1–4 kinetically distinct orientations which were used for the calculations presented in Results.

*Choosing restrained orientations.* We then chose reference orientations for restraining the ligand in the binding site relative to the protein for subsequent free energy calculations. These are defined by picking a specific value in each of six relative protein-ligand degrees of freedom to which to restrain the ligand.<sup>150</sup> These values were chosen as the most probable value of each degree of freedom as determined from histograms

---

<sup>§</sup> <http://folding.stanford.edu/ffamber>

computed during the 1 ns simulations, although in principle this choice is arbitrary.<sup>106, 150</sup>  
The degrees of freedom used are as described previously.<sup>106, 150</sup>

*Binding free energy calculations.* We carried out independent binding free energy calculations for each kinetically distinct orientation. Using the orientational decomposition procedure described previously,<sup>150</sup> we combined the effective binding free energies of each orientation into an overall binding free energy ( $\Delta G_{multiple}^o$ ). We also computed binding free energies that would have resulted had we only considered a single potential bound orientation and neglected symmetry number corrections, as done in docking ( $\Delta G_{single}^o$ ).

Binding free energy calculations were carried out in GROMACS 3.3 (with several crucial bugfixes described previously<sup>150</sup>) using the Bennett acceptance ratio<sup>166, 167</sup> method to estimate free energy differences. To calculate absolute binding free energies, we used a thermodynamic cycle developed previously.<sup>106, 148, 150</sup> In this cycle, we begin with the ligand bound to the protein, then restrain the ligand harmonically to a reference orientation within the binding site. We then annihilate the ligand's partial charges, then decouple its Lennard-Jones interactions with the rest of the system. The final ligand state is equivalent to a non-interacting ligand with no electrostatics, restrained, in vacuum or water. We then analytically calculate the free energy of removing the restraints, and compute the free energy of restoring first the Lennard-Jones and then the electrostatic interactions in water. This entire process forms a thermodynamic cycle that transfers the ligand from the binding site to bulk water in the standard state. If all of the component calculations are converged, this rigorously provides a measurement of the absolute binding free energy,  $\Delta G^o$ , for the forcefield and solvent model used<sup>||</sup>. As part of each of the steps in the cycle, independent free energy calculations were conducted at a number

---

<sup>||</sup> The sum of the components around the cycle is actually the negative of  $\Delta G^o$ .

of intermediate alchemical states (denoted by the parameter  $\lambda$ ) which were the same as those described in our previous work.<sup>150</sup>

Following these binding free energy calculations and the predictions discussed below, we also computed binding free energies for the set of small molecules using AM1-BCC charges. To do this, we computed the free energy of changing AM1-CM2 to AM1-BCC charges in water for each compound, and then repeated the restraining and charging calculations for the compound in the protein for each orientation. Since the Lennard-Jones decoupling is done with compound's electrostatics already turned off, it was unnecessary to repeat these calculations.

*Simulation protocols.* For all of the simulations discussed here (at each  $\lambda$  value), equilibration was performed as follows. First, velocities were assigned from a Maxwell-Boltzmann distribution at 300 K and the the system was subjected to 10 ps of isothermal molecular dynamics. This was followed by 100 ps of isothermal-isobaric dynamics with pressure regulated by the Berendsen weak-coupling scheme<sup>168</sup> as discussed previously.<sup>150</sup> Following this, the simulation cell size was fixed and production simulations were run with isothermal dynamics, using the Langevin integrator for temperature control with a friction coefficient of  $1 \text{ ps}^{-1}$ . Production simulations were 1 ns in length for simulations of the complex (at each alchemical intermediate state, or  $\lambda$  value), and 5 ns for the ligand in water, except where noted otherwise.

All remaining protocols are as discussed previously,<sup>150</sup> with several exceptions: First, PME parameters were modified from those used previously to increase accuracy. Here, we used a PME spline order of 6, a relative tolerance of  $10^{-6}$ , and a Fourier spacing of as close as possible to  $1.0 \text{ \AA}$ . Additionally, we applied a long range van der Waals correction (in addition to the analytical correction employed previously<sup>150</sup>) to correct for the effect of truncating the long-range dispersive interactions at a finite cutoff. These interactions are everywhere attractive, and can contribute significantly to binding free

energies due to the fact that proteins have a higher density of attractive sites than water. While this issue will be discussed in detail elsewhere,<sup>169</sup> the approach used here was, briefly, to run as usual the set of simulations where the ligand Lennard-Jones interactions are decoupled (that is, the ligand-environment Lennard-Jones interactions are turned off). These simulations were then reprocessed with long (24 Å) cutoffs for Lennard-Jones interactions, and the weighted histogram analysis method<sup>170</sup> was used to reweight the data from the simulations conducted with the short cutoff in order to estimate what the decoupling free energy would have been had we run with the longer cutoff. This is a relatively small correction (0.2-0.8 kcal/mol) in the direction of increased binding affinity. This correction would be larger had not an approximate analytical dispersion correction already been included in the original runs by using the GROMACS correction option ENERPRES, and tends to be larger for larger ligands.<sup>169</sup>

To aid convergence of the calculations for benzene, which has a very high symmetry number, we used an approach employed previously<sup>105</sup> and restricted benzene to stay within a single symmetric orientation during our free energy calculations, then simply included the effect of symmetry as a symmetry number correction to the binding free energy.<sup>150</sup>

*Confine-and-Release for Val111.* For some ligands, a valine sidechain (Val111) in the binding cavity is observed experimentally to change rotameric states on ligand binding. This conformational change is not typically sampled (or not well sampled) during our molecular dynamics simulations (discussed in Results). Neglect of this change leads to an underestimate of binding free energies for those ligands when the *apo* protein structure is used, and an overestimate when the *holo* protein structure is used. In the former case, the protein typically remains trapped in a conformation where the valine interferes with ligand binding; in the latter, when the ligand is removed from the binding site, the protein remains kinetically trapped with the sidechain in an unfavorable

orientation, leading to neglect of protein strain energy in the free energy calculation. One way to properly account for the presence of these multiple conformations is to use the “confine-and-release” strategy.<sup>114</sup> The basic idea is to compute the binding free energy of the ligand to the protein, with the protein conformation confined (either kinetically or with artificial restraints) to a particular region of configuration space, then to compute the free energy of releasing the protein from confinement in the bound and unbound states. This provides a rigorous approach for computing binding free energies which include contributions from these conformational changes.<sup>114</sup>

We apply the confine-and-release approach to compute the binding free energies of all the compounds considered here. To do this, we first compute the binding free energy with the valine sidechain kinetically trapped in the orientation from the *apo* structure (which we check by monitoring the dihedral angle throughout all of our simulations). In some cases, the valine actually manages to briefly escape from its kinetic trap at one or several  $\lambda$  values in the alchemical part of the calculation; we discard any simulation snapshots where it had done so from our data analysis in order to apply the confine-and-release approach. Once we have confined binding free energies, computed with the sidechain trapped, we use umbrella sampling and weighted histogram analysis method<sup>170</sup> to compute the potential of mean force for rotating the valine sidechain in the bound and unbound states. From this, we compute the free energy of releasing the protein from confinement, and thus the binding free energies. We present our results with and without the confine-and-release approach, which provides a rigorous way to account for inadequate sampling. Simulation details for the umbrella sampling calculations are as described previously.<sup>114</sup>

Since free energy calculations were conducted with two charge sets, two sets of umbrella sampling calculations for Val111 were carried out for each ligand: one where the ligand had AM1-BCC charges, and one where it had AM1-CM2 charges. This was



important since the details of the ligand electrostatics can influence the free energy landscape associated with this reorientation.

*Predictions.* To predict whether untested compounds would bind, we selected five small molecules predicted by docking (using protocols described previously)<sup>104</sup> to be binders. We followed the same protocols described above: docking to the binding site, retaining a number of different kinetically distinct starting orientations, running separate binding free energy calculations for each of these, and then combining them to get a total binding free energy. We also applied the confine-and-release approach to account for any reorientation of the Val111 sidechain on binding of these molecules. From computed binding free energies, we calculated dissociation constants, and then predict that those compounds with dissociation constants less than 10 mM (the experimental detection threshold for the thermal upshift assay) should bind. These predictions, and those of bound structures below, were made with AM1-CM2 charges, as AM1-BCC charges were only examined later.

Predicting bound structures of these unknown ligands is challenging, since our method is intended to provide an accurate estimate of the binding free energy which includes contributions from a variety of different ligand and protein structures, and neglects any effects of the crystal environment which are present in X-ray structures, as well as other differences. Here, we attempted to identify the dominant bound structure by identifying which kinetically distinct ligand binding orientation contributes most favorably to the total binding free energy; to do this, we calculated occupancy probabilities of the different dominant orientations from their estimated binding free energies. The predicted orientation was the orientation with the highest probability of occupancy. Then, we predicted a bound structure by taking a representative snapshot from a simulation where the ligand remains, without restraints, stably within the region of configuration space corresponding to that orientation. Our predicted bound orientations,

then, were single snapshots from molecular dynamics simulations. For one ligand, we predicted that two orientations would have nearly comparable occupancy probabilities, so we predicted that both would be observed.

After these predictions, we continued retrospective studies and found that the AM1-BCC charge model gave more accurate binding free energies. Therefore, we used the AM1-BCC charge set to make predictions for binding free energies prior to measuring these calorimetrically.

*Binding free energies to a rigid protein.* To compare our methodology more closely with docking, we repeated the free energy calculations using essentially the same protocols, but with the protein held rigid in its prepared starting structure, as is often done in docking. To do so, we used the GROMACS option of defining frozen groups that are held fixed during dynamics. Because there are so many fewer degrees of freedom when the protein is completely rigid, convergence was more rapid, and production simulations required only 100 ps at each  $\lambda$  value. Protocols were otherwise the same, and these calculations used AM1-CM2 charges.

We considered several choices for the rigid protein structure. First, we held the protein rigid in its starting structure (prepared from PDB code 181L). Second, motivated by testing approaches that could easily be applied in docking and scoring, we minimized the entire protein in the presence of each ligand individually in vacuum, and used each of these structures for the appropriate ligand. The RMSD to the starting prepared structure is typically around 0.5 Å with this approach. Finally, we modified this second protocol to allow only residues near the binding site (residues 78, 84, 85, 87, 88, 91, 98, 99, 100, 102, 103, 106, 111, 118, 121, 133, and 153) to move during minimization. With this protocol, changes in structure were very minor (often less than 0.01 Å RMSD (the RMSD reported is for the protein as a whole)).

Minimization protocols were as discussed above and previously,<sup>150</sup> except the order of minimization was reversed (steepest descents followed by limited memory Broyden-Fletcher-Goldfarb-Shanno) and, since minimization was done in vacuum, cutoff electrostatics was used instead of PME, using a cutoff of 11 Å.

*Error analysis.* Calculated uncertainties reported here are one standard deviation of the mean over 40 block bootstrap trials, where the block length is taken to be equal to the autocorrelation time, as described previously.<sup>150</sup>

### **Experimental Methods.**

*Binding detection by upshift of thermal denaturation temperature.* To detect binding, L99A protein was denatured reversibly by temperature in the presence and absence of the putative ligand. Molecules that bind preferentially to the folded cavity-containing protein should stabilize it relative to the *apo* protein, raising its temperature of melting.<sup>113</sup> Thermal denaturation experiments were carried out in a Jasco J-715 spectropolarimeter with a Jasco PTC-348WI Peltier-effect in-cell temperature control device and in-cell stirring. Each compound was screened in its neutral form. 1,2-benzenedithiol was assayed in a pH 3 buffer containing 25mM KCl, 2.9 mM phosphoric acid and 17 mM KH<sub>2</sub>PO<sub>4</sub>. Compounds 1,2-dichlorobenzene and 1-methylpyrrole were screened in a pH 5.4 buffer containing 100 mM sodium chloride, 8.6 mM sodium acetate, and 1.6 mM acetic acid. Compounds thieno[2,3-*c*]pyridine and *n*-methylaniline were screened in a pH 6.8 buffer composed of 50 mM potassium chloride, and 38% (v/v) ethylene glycol. All buffers are as previously described.<sup>113</sup> Thermal denaturation of the protein in the presence of compounds 1,2-dichlorobenzene, thieno[2,3-*c*]pyridine, and *n*-methylaniline were monitored by circular dichroism (CD) between 223 and 234 nm (although the 223 nm wavelength is the ideal wavelength for measuring the helical signal of T4 lysozyme, the higher wavelengths, which were less affected by absorbance from

some of the compounds, can be used to monitor the edge of the helical signal). For 1,2-benzenedithiol and 1-methylpyrrole, which have high absorbance in the far UV region, thermal denaturation was measured by the intensity of the integrated fluorescence emission for all wavelengths above 300 nm, exciting at 290nm. Thermal melts were performed at a temperature ramp rate of 2 K/min. A least-squares fit of the two-state transition model was performed with the program EXAM<sup>129</sup> to calculate  $T_m$  and van't Hoff  $\Delta H$  values for the thermal denaturations. The  $\Delta C_p$  was set to 1.94 kcal mol<sup>-1</sup> K<sup>-1</sup>.

Thermal denaturation of *apo* T4 lysozyme L99A was carried out with 0.02-0.04 mg/ml protein in the same buffer conditions described above. Compounds were included at concentrations as high as 10 mM. Each denaturation experiment was performed at least three times.

*Isothermal titration calorimetry.* Quantitative estimates of association for ligand binding to L99A T4 lysozyme were obtained by ITC using a Microcal VP-ITC calorimeter<sup>171</sup> operated at 10 °C with a reference power of 10  $\mu$ cal/sec, a stirring speed of 300 rpm, and a data collection interval of four minutes per injection. An initial injection of 2  $\mu$ L of ligand was followed by an additional 29 injections of 10  $\mu$ L totaling 292  $\mu$ L. These were added to 0.05 to 0.13 mM protein in the 1.4266 mL sample cell. The concentration of small molecule ligands in the syringe was adjusted such that the final molar ratio of ligand to protein was at least twofold by the end of the titrations. Protein concentrations were determined by molar absorptivity at 280 nm in 0.5 M NaCl, 0.1 M sodium phosphate buffer, pH 6.8. Ligand concentrations were determined by volume of material added to a known volume of buffer. Baseline mixing heats were estimated by injection of ligand into buffer. Reaction heat profiles were fit to the single binding site model using the ITC worksheet of ORIGIN version 7.0 (OriginLab, Northhampton, MA).

*Protein preparation and crystallography.* T4 lysozyme mutant L99A was overexpressed, purified, and crystals grown as described previously.<sup>172</sup> The crystals belong to space group  $P3_22_1$ . Crystals were soaked overnight to four days in crystallization buffer containing as much as 50 mM compound. In addition, drops of neat compound were added to the cover slip surrounding the drop containing the crystal. After soaking, the crystals were cryoprotected with a 50:50 Paraton-N (Hampton Research, Aliso Viejo, CA), mineral oil mix. X-ray data were collected at 110 K with an in-house Raxis IV detector. Reflections were indexed, integrated, and scaled using the HKL package.<sup>77</sup> The complex structures were refined using REFMAC5.<sup>173</sup> For model building and water placement, we used Coot.<sup>132</sup>

**Protein Data Bank accession numbers.** The benzene-bound crystal structure of the T4 lysozyme L99A mutant has been deposited in the Protein Data Bank (PDB) with the accession number 181L. The X-ray crystal structures of 1,2-dichlorobenzene, *n*-methylaniline, and 1-methylpyrrole have been deposited in the PDB with accession numbers 2OTY, 2OTZ, and 2OU0, respectively. The bound structures of indene, *n*-butylbenzene, isobutylbenzene, *p*-xylene, and *o*-xylene have accession numbers 183L, 186L, 184L, 187L, and 188L, respectively.

## **Acknowledgments**

We thank Sarah Boyce for help with protein preparation. We thank Benoît Roux (University of Chicago), Bill Swope (IBM Almaden), Jed Pitera (IBM Almaden), Guha Jayachandran (Stanford University), and Vijay Pande (Stanford University) for helpful discussions, and the reviewers for insightful comments on the manuscript. JDC was supported in part by HHMI and IBM predoctoral fellowships. DLM and KAD acknowledge NIH grant GM63592, Anteon Corporation grant USAF-5408-04-SC-0008, and a UCSF Sandler Award. BKS acknowledges NIH grant GM59957. This work was performed in part with the UCSF QB3 Shared Computing Facility.

## Future Directions

The following section outlines some potential projects or areas of exploration that extend from my work as a graduate student. I can not take full credit for all of the following ideas. Some of them have come from discussions with colleagues. Others have been floating around the lab but have not come to fruition. It is difficult to predict the future especially in scientific discovery. Here, for what it is worth, is my attempt.

**Additional Model Systems and Decoys:** In this thesis, I have described, the hydrophobic and polar cavity sites in T4 lysozyme as well as the charged site in CCP. They are valuable tools for testing computational methods. Additional model systems would allow for further exploration. Brian Matthews' group developed a charged version of the lysozyme cavity by mutating the residue Met102 to a glutamate. This charged site extends the complexity of the lysozyme cavity series. It allows us to address charge-charge and charge-polar interactions in a very simple site. The anionic cavity site in Cytochrome C Peroxidase (CCP W191G) has allowed us to study charge-charge and charge-polar interactions in scoring, but it has the additional complexity of several ordered water molecules and larger protein conformational change. Docking studies against the glutamate mutant in lysozyme are currently underway. Charged ligands will most likely be discovered that bind to the site. No neutral and only a few polar, uncharged ligands bind to the charged CCP site. It would be interesting to see whether this result holds true for the charged lysozyme pocket. For example, benzene binds to the hydrophobic and polar lysozyme sites, but would benzene also bind to the charged site? My guess is that it would not. But how about the polar compounds like phenol?

Interestingly in two of the L99A/M102Q complex crystal structures that I solved in Chapter 2, there was evidence that Phe114 in the cavity rotates and opens a water channel to the surface. This residue should be explored in mutational analysis. I envision

mutating this residue to an alanine or glycine to make a cavity mutant with a small water channel to the surface. This potential mutant would help bridge the gap between the completely buried cavity and solvent exposed sites. This would be a good test system for receptor solvation models.

There have also been attempts to make new CCP mutants. Currently, there is a neutral version of the cavity but few if any known ligands or decoys. Not for lack of testing. A major challenge in testing compounds for the neutral site is finding a convincing assay. Surface Plasmon Resonance (SPR) has shown some initial promise in the lab. This technique should be explored further for all of the model systems. It requires very little protein, and it has the potential to provide binding affinity data. This information is imperative if one would like to fully test the capabilities of high level calculations such as the free energy methods discussed in Chapter 3. There are other mutants of CCP that have not been fully considered. The anionic CCP mutant has a loop that opens the cavity to bulk solvent upon binding larger ligands. The Goodin lab has developed a permanently open pocket by shortening this loop. This site should be pursued for docking studies. As in the potential Phe114 mutant proposed for lysozyme, this provides a link between the very simple cavity sites and more complex drug sites. It would also be a good model system for testing receptor solvation models.

The ligands and especially decoys from these systems provide insight toward our understanding of the energetics of binding. They are an excellent tool for pointing out the strengths and weaknesses of a given computational method. Research with the model systems should continue through a cycle of retrospective and prospective prediction. Any new computational procedure or “improvement” should be critically analyzed with prospective tests. New decoys should be added to the ones that have already been reported. There is an ongoing project which examines free energy methods in the L99A/M102Q sites as well as the charged CCP site that continues work from Chapter 3. These sites will most likely push the free energy methods to its limit. I foresee success in



the polar cavity, but I feel the charged CCP site may introduce too many difficulties: the charged residue, the heme, ordered water, and large conformational change. Either way the results will be informative. Chapter 3 introduced the idea of removing energetic sophistication from the free energy methods to see what components were most important for it's success. In this work we found that holding the receptor rigid was detrimental. Other approximations, such as using an implicit solvent model, should be tested to explore their effects.

While our current experimental methods for testing binding are fairly straight forward they are low throughput. In lieu of testing compounds, we often rely on our ability to guess which compounds will bind to the model sites. We have enough experience to make good educated guesses for many of the hits, but there are several “ambiguous” compounds as reported in Chapter 2. These compounds look significantly different from those that have been tested; therefore, we can not make an educated guess as to whether or not they bind. Again, SPR has some promising potential here. One could screen a whole library of compounds or a whole hit list in just a few days. Though some care should be taken since many of the compounds, especially for the hydrophobic cavity, are not very soluble.

**Algorithm Development:** The work in Chapter 2 and Chapter 3 highlight the importance of receptor flexibility. The inclusion of receptor flexibility and ligand induced fit is a key feature that is left out of computational methods. Several groups such as the Jacobson, Dill, and Abagyan groups have been investigating ways to predict protein flexibility and ligand induced fit. Current molecular dynamics and molecular mechanics protocols can not adequately sample flexible loops and sidechain rotations. Without prior structural evidence, it is difficult to predict even small movements of the receptor. This is a challenging area of research. Homology modeling, loop prediction, and MD methods will likely improve our ability to predict conformational changes especially as computers

become faster. Any structure prediction method must be careful to adequately account for receptor conformational energy change. One also must be aware that these methods may potentially introduce decoy protein conformations and enrich new false positive hits. I foresee a computationally intensive and accurate method such as MD that could be used to generate potential protein structures. This method may be able to adequately account for the energy differences between different conformations. The starting structures and their energies could then be used by docking methods.

A former graduate student from the lab, Binqing Wei, developed a receptor flexibility algorithm for docking that relies on previously determined structural data.<sup>65</sup> It can not predict motions for which we have no structural information, though it can be used in cases where several structures are available from the PDB or computationally generated. Unfortunately, his algorithm has not been incorporated into our standard docking code. It should be a priority to add his method into the current code. I doubt it would be a difficult project, and the potential benefits far outweigh the time it would take to include it. The flexible docking method could be tested in the CCP pockets for which there are large loop movements. This method also has the potential to test the flexibility of ordered waters and predict those required upon ligand binding.

In docking, we generate many poses but only keep the one that scores the best. DOCK4 and later codes from the Kuntz's group have the ability to save clusters of ligand poses. This should be incorporated into our own version of the code. DOCK often scores an incorrect pose for a ligand better than native-like poses which are not far down the list. These other poses are useful from a rescoring standpoint. Many ligands have multiple binding modes that contribute to their total binding energies as discussed in Chapter 3. Clusters of ligand poses obtained from multiple docking positions could also be used to estimate the contributions of translational and rotational entropy during docking as described by Ruvinsky et al.<sup>174</sup>

Solvation models are another key problem for the computational field. Explicit solvent models are the best and most accurate approaches. They are computationally very slow and still have difficulty estimating hydration free energies of polar and charged ligands. Faster docking methods will most likely rely on implicit solvent models for a long time to come. The implicit methods that we have should be improved. Most use an internal dielectric of one, and their partial charges have been fit using this methodology. This is fundamentally wrong and should be corrected.

## References

1. Kauffman, G. B. & Chooljian, S. H. (2001). Friedrich Wohler (1800-1882), on the Bicentennial of His Birth. *Chem. Educator* 6, 121-133.
2. Barnett, J. A. & Lichtenthaler, F. W. (2001). A history of research on yeasts 3: Emil Fischer, Eduard Buchner and their contemporaries, 1880-1900. *Yeast* 18, 363-388.
3. Meyer, E. F. (1995). Emil Fischer: then and now. *Pharmaceutica Acta Helvetiae* 69, 177-183.
4. Lucretius. (1995). *On the Nature of Things: De rerum natura*. Trans. Esolen, A. M., The Johns Hopkins University Press, Baltimore.
5. Jones, R. (2001). Nonsteroidal anti-inflammatory drug prescribing: past, present, and future. 110, S4-S7.
6. Vane, J. R. (1971). Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs. *Nature New Biology* 23, 232-235.
7. Picot, D., Loll, P. J. & Garavito, R. M. (1994). The X-ray crystal structure of the membrane protein prostaglandin H<sub>2</sub> synthase-1. *Nature* 367, 243-249.
8. Mould, R. F. (1995). The early history of X-ray diagnosis with emphasis on the contributions of physics 1895-1915. *Physics in Medicine and Biology* 40, 1741-1787.
9. Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G. & North, A. C. T. (1960). Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature* 185, 416-422.
10. Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. C. (1960). Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å Resolution. *Nature* 185, 422-427.

11. RCSB Protein Data Bank website: <http://www.rcsb.org>.
12. Beddell, C. R., Goodford, P. J., Norrington, F. E., Wilkinson, S. & Wootton, R. (1976). Compounds designed to fit a site of known structure in human haemoglobin. *British Journal of Pharmacology* 57, 201-209.
13. Cushman, D. W., Cheung, H. S., Sabo, E. F. & Ondetti, M. A. (1977). Design of potent competitive inhibitors of angiotensin-converting enzyme. Carboxyalkanoyl and mercaptoalkanoyl amino acids. *Biochemistry* 16, 5484-5491.
14. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *J Mol Biol* 161, 269-88.
15. Blaney, J. M., Jorgensen, E. C., Connolly, M. L., Ferrin, T. E., Langridge, R., Oatley, S. J., Burrige, J. M. & Blake, C. C. F. (1982). Computer graphics in drug design: molecular modeling of thyroid hormone-prealbumin interactions. *J. Med. Chem.* 25, 785-790.
16. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P. (1984). A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J Am Chem Soc* 106, 765-784.
17. Meng, E. C., Shoichet, B. K. & Kuntz, I. D. (1992). Automated Docking with Grid-Based Energy Evaluation. *J. Comput. Chem.* 13, 505-524.
18. Shoichet, B. K., Leach, A. R. & Kuntz, I. D. (1999). Ligand solvation in molecular docking. *Proteins: Structure, Function, and Genetics* 34, 4-16.
19. Jorgensen, W. L., Maxwell, D. S. & TiradoRives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118, 11225-11236.
20. Gohlke, H., Hendlich, M. & Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* 295, 337-356.
21. Wang, R., Lu, Y. & Wang, S. (2003). Comparative evaluation of 11 scoring functions for molecular docking. *J Med Chem* 46, 2287-303.

22. Muegge, I. & Martin, Y. C. (1999). A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *J. Med. Chem.* 42, 791-804.
23. Powers, R. A., Morandi, F. & Shoichet, B. K. (2002). Structure-Based Discovery of a Novel, Noncovalent Inhibitor of AmpC beta-Lactamase. *Structure* 10, 1013-1023.
24. Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* 261, 470-489.
25. Verkhivker, G. M., Bouzida, D., Gehlhaar, D. K., Rejto, P. A., Arthurs, S., Colson, A. B., Freer, S. T., Larson, V., Luty, B. A., Marrone, T. & Rose, P. W. (2000). Deciphering common failures in molecular docking of ligand-protein complexes. *J Comput Aided Mol Des* 14, 731-51.
26. Ishchenko, A. V. & Shakhnovich, E. I. (2002). Small Molecule Growth 2001 (SMoG2001): An Improved Knowledge-Based Scoring Function for Protein-Ligand Interactions. *J. Med. Chem.* 45, 2770-2780.
27. Park, B. & Levitt, M. (1996). Energy Functions that Discriminate X-ray and Near-native Folds from Well-constructed Decoys. *J Mol Biol* 258, 367-392.
28. Samudrala, R. & Levitt, M. (2000). Decoys 'R' Us: a database of incorrect conformations to improve protein structure prediction. *Protein Sci* 9, 1399-1401.
29. Seok, C., Rosen, J. B., Chodera, J. D. & Dill, K. A. (2003). MOPED: Method for optimizing physical energy parameters using decoys. *J. Comput. Chem.* 24, 89-97.
30. Camacho, C. J., Gatchell, D. W., Kimura, S. R. & Vajda, S. (2000). Scoring docked conformations generated by rigid-body protein-protein docking. *Proteins: Structure, Function, and Genetics* 40, 525-537.
31. Burkhard, P., Hommel, U., Sanner, M. & Walkinshaw, M. D. (1999). The discovery of steroids and other novel FKBP inhibitors using a molecular docking program. *J. Mol. Biol.* 287, 853-858.
32. Enyedy, I. J., Ling, Y., Nacro, K., Tomita, Y., Wu, X., Cao, Y., Guo, R., Li, B., Zhu, X., Huang, Y., Long, Y. Q., Roller, P. P., Yang, D. & Wang, S. (2001).

Discovery of small-molecule inhibitors of Bcl-2 through structure-based computer screening. *J Med Chem* 44, 4313-24.

33. Grädler, U., Gerber, H. D., Goodenough-Lashua, D. M., Garcia, G. A., Ficner, R., Reuter, K., Stubbs, M. T. & Klebe, G. (2001). A New Target for Shigellosis: Rational Design and Crystallographic Studies of Inhibitors of tRNA-guanine Transglycosylase. *J Mol Biol* 306, 455-467.
34. Honma, T., Hayashi, K., Aoyama, T., Hashimoto, N., Machida, T., Fukasawa, K., Iwama, T., Ikeura, C., Ikuta, M., Suzuki-Takahashi, I., Iwasawa, Y., Hayama, T., Nishimura, S. & Morishima, H. (2001). Structure-Based Generation of a New Class of Potent Cdk4 Inhibitors: New de Novo Design Strategy and Library Design. *J. Med. Chem.* 44, 4615-4627.
35. Liebeschuetz, J. W., Jones, S. D., Morgan, P. J., Murray, C. W., Rimmer, A. D., Roscoe, J. M. E., Waszkowycz, B., Welsh, P. M., Wylie, W. A., Young, S. C., Martin, H., Mahler, J., Brady, L. & Wilkinson, K. (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, 1221-1232.
36. Ealick, S. E., Babu, Y. S., Bugg, C. E., Erion, M. D., Guida, W. C., Montgomery, J. A. & Secrist Ja, III. (1991). Application of Crystallographic and Modeling Methods in the Design of Purine Nucleoside Phosphorylase Inhibitors. *PNAS* 88, 11540-11544.
37. Iwata, Y., Arisawa, M., Hamada, R., Kita, Y., Mizutani, M. Y., Tomioka, N., Itai, A. & Miyamoto, S. (2001). Discovery of Novel Aldose Reductase Inhibitors Using a Protein Structure-Based Approach: 3D-Database Search Followed by Design and Synthesis. *J. Med. Chem.* 44, 1718-1728.
38. Paiva, A. M., Vanderwall, D. E., Blanchard, J. S., Kozarich, J. W., Williamson, J. M. & Kelly, T. M. (2001). Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of Mycobacterium tuberculosis. *Biochim Biophys Acta* 1545, 67-77.
39. DesJarlais, R. L., Seibel, G. L., Kuntz, I. D., Furth, P. S., Alvarez, J. C., Ortiz de Montellano, P. R., DeCamp, D. L., Babe, L. M. & Craik, C. S. (1990). Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease. *Proc Natl Acad Sci USA* 87, 6644-8.

40. Boehm, H. J., Boehringer, M., Bur, D., Gmuender, H., Huber, W., Klaus, W., Kostrewa, D., Kuehne, H., Luebbers, T., Meunier-Keller, N. & Mueller, F. (2000). Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *J. Med. Chem.* 43, 2664-2674.
41. Grzybowski, B. A., Ishchenko, A. V., Kim, C.-Y., Topalov, G., Chapman, R., Christianson, D. W., Whitesides, G. M. & Shakhnovich, E. I. (2002). Combinatorial computational method gives new picomolar ligands for a known enzyme. *PNAS* 99, 1270-1273.
42. Rao, M. S. & Olson, A. J. (1999). Modelling of Factor Xa-inhibitor complexes: a computational flexible docking approach. *Proteins: Structure, Function, and Genetics* 34, 173-183.
43. Vangrevelinghe, E., Zimmermann, K., Schoepfer, J., Portmann, R., Fabbro, D. & Furet, P. (2003). Discovery of a potent and selective protein kinase CK2 inhibitor by high-throughput docking. *J Med Chem* 46, 2656-62.
44. Verras, A., Kuntz, I. D. & Ortiz de Montellano, P. R. (2004). Computer-assisted design of selective imidazole inhibitors for cytochrome p450 enzymes. *J Med Chem* 47, 3572-9.
45. Doman, T. N., McGovern, S. L., Witherbee, B. J., Kasten, T. P., Kurumbail, R., Stallings, W. C., Connolly, D. T. & Shoichet, B. K. (2002). Molecular Docking and High-Throughput Screening for Novel Inhibitors of Protein Tyrosine Phosphatase-1B. *J. Med. Chem.* accepted with revisions.
46. Abagyan, R. & Totrov, M. (2001). High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* 5, 375-382.
47. Perola, E., Walters, W. P. & Charifson, P. S. (2004). A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* 56, 235-49.
48. Mirny, L. A. & Shakhnovich, E. I. (1996). How to Derive a Protein Folding Potential? A New Approach to an Old Problem. *J Mol Biol* 264, 1164-1179.



49. Simons, K. T., Bonneau, R., Ruczinski, I. & Baker, D. (1999). Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins: Structure, Function, and Genetics* 37, 171-176.
50. Chiu, T. L. & Goldstein, R. A. (2000). How to generate improved potentials for protein tertiary structure prediction: A lattice model study. *Proteins: Structure, Function, and Genetics* 41, 157-163.
51. Felts, A. K., Gallicchio, E., Wallqvist, A. & Levy, R. M. (2002). Distinguishing native conformations of proteins from decoys with an effective free energy estimator based on the OPLS all-atom force field and the surface generalized born solvent model. *Proteins: Structure, Function, and Genetics* 48, 404-422.
52. Tobi, D. & Elber, R. (2000). Distance-dependent, pair potential for protein folding: Results from linear optimization. *Proteins: Structure, Function, and Genetics* 41, 40-46.
53. Murphy, J., Gatchell, D. W., Prasad, J. C. & Vajda, S. (2003). Combination of scoring functions improves discrimination in protein-protein docking. *Proteins: Structure, Function, and Genetics* 53, 840-854.
54. Sternberg, M. J., Gabb, H. A., Jackson, R. M. & Moont, G. (2000). Protein-protein docking. Generation and filtering of complexes. *Methods Mol. Biol.* 143, 399-415.
55. Gabb, H. A., Jackson, R. M. & Sternberg, M. J. (1997). Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J Mol Biol* 272, 106-120.
56. Holm, L. & Sander, C. (1992). Evaluation of protein models by atomic solvation preference. *J Mol Biol* 225, 93-105.
57. Mandell, J. G., Robers, V. A., Pique, M. E., Kotlovyyi, V., Mitchell, J. C., Nelson, E., Tsigelny, I. & Ten Eyck, L. F. (2001). Protein docking using continuum electrostatics and geometric fit. *Protein Eng* 14, 105-113.
58. Eriksson, A. E., Baase, W. A., Wozniak, J. A. & Matthews, B. W. (1992). A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature* 355, 371-3.

59. Wei, B. Q., Baase, W. A., Weaver, L. H., Matthews, B. W. & Shoichet, B. K. (2002). A model binding site for testing scoring functions in molecular docking. *J Mol Biol* 322, 339-355.
60. Stahl, M. & Rarey, M. (2001). Detailed analysis of scoring functions for virtual screening. *J Med Chem* 44, 1035-42.
61. Gohlke, H. & Klebe, G. (2002). Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem., Int. Ed. Engl.* 41, 2644-2676.
62. Ferrara, P., Gohlke, H., Price, D. J., Klebe, G. & Brooks, C. L., 3rd. (2004). Assessing scoring functions for protein-ligand interactions. *J Med Chem* 47, 3032-47.
63. Morton, A., Baase, W. A. & Matthews, B. W. (1995). Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochemistry* 34, 8564-75.
64. Su, A. I., Lorber, D. M., Weston, G. S., Baase, W. A., Matthews, B. W. & Shoichet, B. K. (2001). Docking molecules by families to increase the diversity of hits in database screens: Computational strategy and experimental evaluation. *Proteins* 42, 279-293.
65. Wei, B. Q., Weaver, L. H., Ferrari, A. M., Matthews, B. W. & Shoichet, B. K. (2004). Testing a Flexible-receptor Docking Algorithm in a Model Binding Site. *J Mol Biol* 337, 1161-82.
66. Tondi, D., Morandi, F., Bonnet, R., Costi, M. P. & Shoichet, B. K. (2005). Structure-Based Optimization of a Non- $\beta$ -lactam Lead Results in Inhibitors That Do Not Up-Regulate  $\beta$ -Lactamase Expression in Cell Culture. *J. Am. Chem. Soc.* 127, 4632-4639.
67. Lorber, D. M. & Shoichet, B. K. (1998). Flexible ligand docking using conformational ensembles. *Protein Sci* 7, 938-50.
68. McGovern, S. L. & Shoichet, B. K. (2003). Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J Med Chem* 46, 2895-907.

69. Shoichet, B. K. & Kuntz, I. D. (1993). Matching chemistry and shape in molecular docking. *Protein Eng* 6, 723-32.
70. Gschwend, D. A. & Kuntz, I. D. (1996). Orientational sampling and rigid-body minimization in molecular docking revisited: On-the-fly optimization and degeneracy removal. *J. Comput.-Aided Mol. Des.* 10, 123-132.
71. Li, J. B., Zhu, T. H., Cramer, C. J. & Truhlar, D. G. (1998). New class IV charge model for extracting accurate partial charges from wave functions. *J. Phys. Chem. A* 102, 1820-1831.
72. Chambers, C. C., Hawkins, G. D., Cramer, C. J. & Truhlar, D. G. (1996). Model for aqueous solvation based on class IV atomic charges and first solvation shell effects. *J. Phys. Chem.* 100, 16385-16398.
73. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605-1612.
74. McGovern, S. L., Caselli, E., Grigorieff, N. & Shoichet, B. K. (2002). A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J Med Chem* 45, 1712-22.
75. McGovern, S. L., Helfand, B. T., Feng, B. & Shoichet, B. K. (2003). A Specific Mechanism of Nonspecific Inhibition. *J. Med. Chem.* 46, 4265-4272.
76. Lipscomb, L. A., Gassner, N. C., Snow, S. D., Eldridge, A. M., Baase, W. A., Drew, D. L. & Matthews, B. W. (1998). Context-dependent protein stabilization by methionine-to-leucine substitution shown in T4 lysozyme. *Protein Sci.* 7, 765-773.
77. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In *Method Enzymol.* (Carter, C. W. J. & Sweet, R. M., eds.), Vol. 276, pp. 307-326. Academic Press, New York.
78. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54, 905-921.

79. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C. & Mainz, D. T. (2006). Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* 49, 6177-6196.
80. Wang, J., Kang, X., Kuntz, I. D. & Kollman, P. A. (2005). Hierarchical database screenings for HIV-1 reverse transcriptase using a pharmacophore model, rigid docking, solvation docking, and MM-PB/SA. *J Med Chem* 48, 2432-44.
81. Jorgensen, W. L., Ruiz-Caro, J., Tirado-Rives, J., Basavapathruni, A., Anderson, K. S. & Hamilton, A. D. (2006). Computer-aided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* 16, 663-667.
82. Claussen, H., Buning, C., Rarey, M. & Lengauer, T. (2001). FlexE: efficient molecular docking considering protein structure variations. *J Mol Biol* 308, 377-95.
83. Cavasotto, C. N. & Abagyan, R. A. (2004). Protein flexibility in ligand docking and virtual screening to protein kinases. *J Mol Biol* 337, 209-25.
84. Chang, C. E. & Gilson, M. K. (2004). Free energy, entropy, and induced fit in host-guest recognition: calculations with the second-generation mining minima algorithm. *J Am Chem Soc* 126, 13156-64.
85. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A. & Farid, R. (2006). Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* 49, 534-553.
86. Evers, A. & Klebe, G. (2004). Successful Virtual Screening for a Submicromolar Antagonist of the Neurokinin-1 Receptor Based on a Ligand-Supported Homology Model. *J. Med. Chem.* 47, 5381-5392.
87. Li, S., Gao, J., Satoh, T., Friedman, Thea M., Edling, Andrea E., Koch, U., Choksi, S., Han, X., Korngold, R. & Huang, Z. (1997). A computer screening approach to immunoglobulin superfamily structures and interactions: Discovery of small non-peptidic CD4 inhibitors as novel immunotherapeutics. *PNAS* 94, 73-78.
88. Huang, N., Nagarsekar, A., Xia, G., Hayashi, J. & MacKerell, A. D., Jr. (2004). Identification of non-phosphate-containing small molecular weight inhibitors of

the tyrosine kinase p56 Lck SH2 domain via in silico screening against the pY + 3 binding site. *J Med Chem* 47, 3502-11.

89. Song, H., Wang, R., Wang, S. & Lin, J. (2005). A low-molecular-weight compound discovered through virtual database screening inhibits Stat3 function in breast cancer cells. *PNAS* 102, 4700-4705.
90. Lyne, P. D., Kenny, P. W., Cosgrove, D. A., Deng, C., Zabudoff, S., Wendoloski, J. J. & Ashwell, S. (2004). Identification of compounds with nanomolar binding affinity for checkpoint kinase-1 using knowledge-based virtual screening. *J Med Chem* 47, 1962-8.
91. Charifson, P. S., Corkery, J. J., Murcko, M. A. & Walters, W. P. (1999). Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* 42, 5100-5109.
92. Bissantz, C., Folkers, G. & Rognan, D. (2000). Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* 43, 4759-4767.
93. Gohlke, H. & Klebe, G. (2001). Statistical potentials and scoring functions applied to protein-ligand binding. *Curr Opin Struct Biol* 11, 231-5.
94. Wang, R. X. & Wang, S. M. (2001). How does consensus scoring work for virtual library screening? An idealized computer experiment. *J. Chem. Inf. Comput. Sci.* 41, 1422-1426.
95. Kalyanaraman, C., Bernacki, K. & Jacobson, M. P. (2005). Virtual screening against highly charged active sites: identifying substrates of alpha-beta barrel enzymes. *Biochemistry* 44, 2059-71.
96. Huang, N., Kalyanaraman, C., Irwin, J. J. & Jacobson, M. P. (2006). Physics-Based Scoring of Protein-Ligand Complexes: Enrichment of Known Inhibitors in Large-Scale Virtual Screening. *J. Chem. Inf. Model.* 46, 243-253.
97. Emanuele Perola. (2006). Minimizing false positives in kinase virtual screens. *Proteins: Structure, Function, and Bioinformatics* 64, 422-435.

98. Lee, M. R. & Sun, Y. (2007). Improving Docking Accuracy through Molecular Mechanics Generalized Born Optimization and Scoring. *J. Chem. Theory Comput.* 3, 1106-1119.
99. Lyne, P. D., Lamb, M. L. & Saeh, J. C. (2006). Accurate Prediction of the Relative Potencies of Members of a Series of Kinase Inhibitors Using Molecular Docking and MM-GBSA Scoring. *J. Med. Chem.* 49, 4805-4808.
100. Mobley, D. L., Graves, A. P., Chodera, J. D., McReynolds, A. C., Shoichet, B. K. & Dill, K. A. (2007). Predicting Absolute Ligand Binding Free Energies to a Simple Model Site. *J Mol Biol* 371, 1118-1134.
101. Fitzgerald, M. M., Churchill, M. J., McRee, D. E. & Goodin, D. B. (1994). Small molecule binding to an artificially created cavity at the active site of cytochrome c peroxidase. *Biochemistry* 33, 3807-18.
102. Musah, R. A., Jensen, G. M., Bunte, S. W., Rosenfeld, R. J. & Goodin, D. B. (2002). Artificial protein cavities as specific ligand-binding templates: characterization of an engineered heterocyclic cation-binding site that preserves the evolved specificity of the parent protein. *J Mol Biol* 315, 845-57.
103. Brenk, R., Vetter, S. W., Boyce, S. E., Goodin, D. B. & Shoichet, B. K. (2006). Probing molecular docking in a charged model binding site. *J Mol Biol* 357, 1449-70.
104. Graves, A. P., Brenk, R. & Shoichet, B. K. (2005). Decoys for docking. *J Med Chem* 48, 3714-28.
105. Deng, Y. & Roux, B. (2006). Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant. *J. Chem. Theory Comput.* 2, 1255-1273.
106. Boresch, S., Tettinger, F., Leitgeb, M. & Karplus, M. (2003). Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B* 107, 9535-9551.
107. Machicado, C., Lopez-Llano, J., Cuesta-Lopez, S., Bueno, M. & Sancho, J. (2005). Design of ligand binding to an engineered protein cavity using virtual screening and thermal up-shift evaluation. *J. Comput.-Aided Mol. Des.* 19, 421-443.

108. Rosenfeld, R., Goodsell, D., Musah, R., Garret, M., Goodin, D. & Olson, A. (2003). Automated docking of ligands to an artificial active site: augmenting crystallographic analysis with computer modeling. *J Comput Aided Mol Des*.
109. Elowe, N. H., Blanchard, J. E., Cechetto, J. D. & Brown, E. D. (2005). Experimental Screening of Dihydrofolate Reductase Yields a "Test Set" of 50,000 Small Molecules for a Computational Data-Mining and Docking Competition. *J Biomol Screen* 10, 653-657.
110. DeLano, W. L. (2002). The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA.
111. Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J., Honig, B., Shaw, D. E. & Friesner, R. A. (2004). A hierarchical approach to all-atom protein loop prediction. *Proteins* 55, 351-67.
112. Jacobson, M. P., Kaminski, G. A., Friesner, R. A. & Rapp, C. S. (2002). Force Field Validation Using Protein Side Chain Prediction. *J. Phys. Chem. B* 106, 11673-11680.
113. Morton, A. & Matthews, B. W. (1995). Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: linkage of dynamics and structural plasticity. *Biochemistry* 34, 8576-88.
114. Mobley, D. L., Chodera, J. D. & Dill, K. A. (2007). Confine-and-Release Method: Obtaining Correct Binding Free Energies in the Presence of Protein Conformational Change. *J. Chem. Theory Comput.* 3, 1231-1235.
115. Song, L., Kalyanaraman, C., Fedorov, A. A., Fedorov, E. V., Glasner, M. E., Brown, S., Imker, H. J., Babbitt, P. C., Almo, S. C., Jacobson, M. P. & Gerlt, J. A. (2007). Prediction and assignment of function for a divergent N-succinyl amino acid racemase. *Nature Chemical Biology* 3, 486-491.
116. Leach, A. R., Shoichet, B. K. & Peishoff, C. E. (2006). Prediction of Protein-Ligand Interactions. Docking and Scoring: Successes and Gaps. *J. Med. Chem.* 49, 5851-5855.
117. Banks, J. L., Beard, H. S., Cao, Y., Cho, A. E., Damm, W., Farid, R., Felts, A. K., Halgren, T. A., Mainz, D. T., Maple, J. R., Murphy, R., Philipp, D. M., Repasky, M. P., Zhang, L. Y., Berne, B. J., Friesner, R. A., Gallicchio, E. & Levy, R. M.

- (2005). Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J Comput Chem* 26, 1752-80.
118. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. (2001). Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* 105, 6474-6487.
  119. Gallicchio, E., Zhang, L. Y. & Levy, R. M. (2002). The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. *J. Comput. Chem.* 23, 517-529.
  120. Zhu, K., Shirts, M. R., Friesner, R. A. & Jacobson, M. P. (2007). Multiscale Optimization of a Truncated Newton Minimization Algorithm and Application to Proteins and Protein-Ligand Complexes. *J. Chem. Theory Comput.* 3, 640-648.
  121. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* 25, 247-260.
  122. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. (2004). Development and testing of a general amber force field. *J Comput Chem* 25, 1157-74.
  123. Gasteiger, J. & Marsili, M. (1980). Iterative partial equalization of orbital electronegativity: a rapid access to atomic charges. *Tetrahedron* 36, 3219-3288.
  124. Jakalian, A., Jack, D. B. & Bayly, C. I. (2002). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* 23, 1623-1641.
  125. Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. (1993). A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J. Phys. Chem.* 97, 10269-10280.
  126. Giammona, D. A. (1984). An examination of conformational flexibility in porphyrins and bulky-ligand binding in myoglobin. Ph.D. Dissertation, University of California, Davis.



127. Onufriev, A., Bashford, D. & Case, D. A. (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* 55, 383-394.
128. Weiser, J., Shenkin, P. S. & Still, W. C. (1999). Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* 20, 217-230.
129. Kirchhoff, W. (1993). EXAM: A two-state thermodynamic analysis program, NIST, Gaithersburg, Maryland.
130. Schuttelkopf, A. W. & van Aalten, D. M. F. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica Section D-Biological Crystallography* 60, 1355-1363.
131. Bailey, S. (1994). The Ccp4 Suite - Programs for Protein Crystallography. *Acta Crystallographica Section D-Biological Crystallography Biol Crystallogr* 50, 760-763.
132. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-32.
133. Halperin, I., Ma, B., Wolfson, H. & Nussinov, R. (2002). Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Structure, Function, and Genetics* 47, 409-443.
134. Huang, N., Kalyanaraman, C., Bernacki, K. & Jacobson, M. P. (2006). Molecular mechanics methods for predicting protein-ligand binding. *Phys. Chem. Chem. Phys.* 8, 5166-5177.
135. Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A. & Case, D. A. (1998). Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *J. Am. Chem. Soc.* 120, 9401-9409.
136. Kuhn, B. & Kollman, P. A. (2000). Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models. *J. Med. Chem.* 43, 3786-3791.

137. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A. & Cheatham, T. E., 3rd. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 33, 889-97.
138. Allen, T. W., Andersen, O. S. & Roux, B. (2004). On the Importance of Atomic Fluctuations, Protein Flexibility, and Solvent in Ion Permeation. *J. Gen. Physiol.* 124, 679-690.
139. Lee, M. S. & Olson, M. A. (2006). Calculation of Absolute Protein-Ligand Binding Affinity Using Path and Endpoint Approaches. *Biophys. J.* 90, 864-877.
140. Pearlman, D. A. (2005). Evaluating the Molecular Mechanics Poisson-Boltzmann Surface Area Free Energy Method Using a Congeneric Series of Ligands to p38 MAP Kinase. *J. Med. Chem.* 48, 7796-7807.
141. Steinbrecher, T., Case, D. A. & Labahn, A. (2006). A Multistep Approach to Structure-Based Drug Design: Studying Ligand Binding at the Human Neutrophil Elastase. *J. Med. Chem.* 49, 1837-1844.
142. Woo, H.-J. & Roux, B. (2005). Chemical Theory and Computation Special Feature: Calculation of absolute protein-ligand binding free energy from computer simulations. *PNAS* 102, 6825-6830.
143. Zhang, D., Gullingsrud, J. & McCammon, J. A. (2006). Potentials of Mean Force for Acetylcholine Unbinding from the Alpha7 Nicotinic Acetylcholine Receptor Ligand-Binding Domain. *J. Am. Chem. Soc.* 128, 3019-3026.
144. Rodinger, T. & Pomès, R. (2005). Enhancing the accuracy, the efficiency and the scope of free energy simulations. *Curr Opin Struct Biol* 15, 164-170.
145. Kofke, D. A. (2005). Free energy methods in molecular simulation. *Fluid Phase Equilib* 228-229, 41-48.
146. Shirts, M. R., Mobley, D. L. & Chodera, J. D. Alchemical free energy calculations: ready for prime time? *Ann. Rep. Comp. Chem.* In the press.

147. Roux, B., Nina, M., Pomès, C. & Smith, J. C. (1996). Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: a molecular dynamics free energy perturbation study. *Biophys. J.* 71, 670-681.
148. Gilson, M. K., Given, J. A., Bush, B. L. & McCammon, J. A. (1997). The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* 72, 1047-1069.
149. Hermans, J. & Wang, L. (1997). Inclusion of Loss of Translational and Rotational Freedom in Theoretical Estimates of Free Energies of Binding. Application to a Complex of Benzene and Mutant T4 Lysozyme. *J. Am. Chem. Soc.* 119, 2707-2714.
150. Mobley, D. L., Chodera, J. D. & Dill, K. A. (2006). On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *The Journal of Chemical Physics* 125, 084902.
151. Wang, J., Deng, Y. & Roux, B. (2006). Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraining Potentials. *Biophys. J.* 91, 2798-2814.
152. Fujitani, H., Tanida, Y., Ito, M., Jayachandran, G., Snow, C. D., Shirts, M. R., Sorin, E. J. & Pande, V. S. (2005). Direct calculation of the binding free energies of FKBP ligands. *The Journal of Chemical Physics* 123, 084108.
153. van den Bosch, M., Swart, M., Snijders, J. G., Berendsen, H. J. C., Mark, A. E., Oostenbrink, C., van Gunsteren, W. F. & Canters, G. W. (2005). Calculation of the Redox Potential of the Protein Azurin and Some Mutants. *ChemBioChem* 6, 738-746.
154. Zhou, Y., Oostenbrink, C., Jongejan, A., van Gunsteren, W. F., Hagen, W. R., de Leeuw, S. W. & Jongejan, J. A. (2006). Computational study of ground-state chiral induction in small peptides: Comparison of the relative stability of selected amino acid dimers and oligomers in homochiral and heterochiral combinations. *J. Comput. Chem.* 27, 857-867.
155. Shirts, M. R. (2005). Calculating precise and accurate free energies in biomolecular systems. Ph.D. dissertation, Stanford University.

156. Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P. & Matthews, B. W. (1992). Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* 255, 178-183.
157. Eriksson, A. E., Baase, W. A. & Matthews, B. W. (1993). Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4 lysozyme have different structural and thermodynamic consequences. *J Mol Biol* 229, 747-69.
158. Mann, G. & Hermans, J. (2000). Modeling protein-small molecule interactions: structure and thermodynamics of noble gases binding in a cavity in mutant phage T4 Lysozyme L99A. *J Mol Biol* 302, 979-989.
159. Mobley, D. L., Dumont, E., Chodera, J. D. & Dill, K. A. (2007). Comparison of Charge Models for Fixed-Charge Force Fields: Small-Molecule Hydration Free Energies in Explicit Solvent. *J. Phys. Chem. B* 111, 2242-2254.
160. Turnbull, W. B. & Daranas, A. H. (2003). On the Value of  $c$ : Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? *J. Am. Chem. Soc.* 125, 14859-14866.
161. Lindahl, E., Hess, B. & van der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* 7, 306-317.
162. van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *J. Comput. Chem.* 26, 1701-1718.
163. Sorin, E. J. & Pande, V. S. (2005). Exploring the Helix-Coil Transition via All-Atom Equilibrium Ensemble Simulations. *Biophys. J.* 88, 2472-2493.
164. Kollman, P. A. (1996). Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules. *Acc. Chem. Res.* 29, 461-469.
165. Jakalian, A., Bush, B. L., Jack, D. B. & Bayly, C. I. (2000). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* 21, 132-146.

166. Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Chem.* 22, 245-268.
167. Shirts, M. R. & Pande, V. S. (2005). Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *The Journal of Chemical Physics* 122, 144107-1–144107-16.
168. Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. v., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* 81, 3684-3690.
169. Shirts, M. R., Mobley, D. L., Chodera, J. D. & Pande, V. S. Accurate and efficient corrections for missing dispersion interactions in molecular simulations. *J Chem Phys B* Submitted.
170. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules: I. The method. *J. Comput. Chem.* 13, 1011-1021.
171. Plotnikov, V. V., Brandts, J. M., Lin, L. N. & Brandts, J. F. (1997). A new ultrasensitive scanning calorimeter. *Anal. Biochem.* 250, 237-244.
172. Poteete, A. R., Sun, D. P., Nicholson, H. & Matthews, B. W. (1991). Second-site revertants of an inactive T4 lysozyme mutant restore activity by restructuring the active site cleft. *Biochemistry* 30, 1425-32.
173. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D* 53, 240-255.
174. Ruvinsky, A. M. & Kozintsev, A. V. (2005). New and fast statistical-thermodynamic method for computation of protein-ligand binding entropy substantially improves docking accuracy. *J. Comput. Chem.* 26, 1089-1095.

## **Appendix A.**

### **Supporting Material**

#### **Decoys for Docking**

Alan P. Graves, Ruth Brenk, Brian K. Shoichet

**Contents: Table S1-S7, Figure S1-S7**

**Table S1.** Ligand bound protein structures from the PDB used to test for geometric decoys.

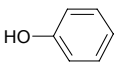
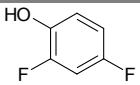
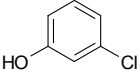
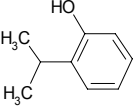
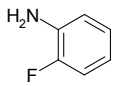
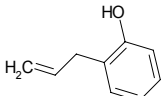
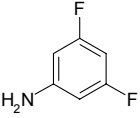
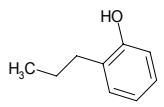
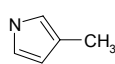
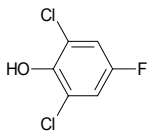
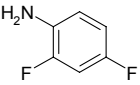
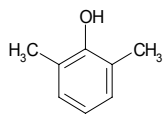
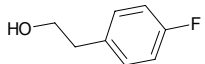
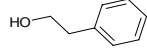
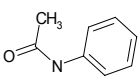
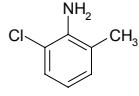
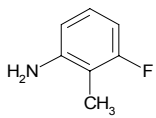
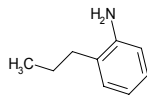
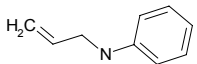
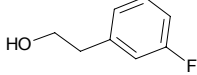
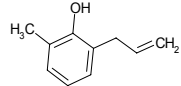
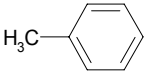
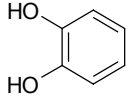
DHFR			
1AOE.pdb (GW3)	1BOZ.pdb (PRD)	1BZF.pdb (TMQ)	1D8R.pdb (TABAB)
1DG5.pdb (TOP)	1DG7.pdb (WRB)	1DIS.pdb (BDM)	1DLR.pdb (MXA)
1DR3.pdb (BIO)	1DYH.pdb (DZF)	1E26.pdb (GPB)	1HFP.pdb (MOT)
1IA1.pdb (TQ3)	1IA2.pdb (TQ4)	1IA3.pdb (TQ5)	1IA4.pdb (TQ6)
1RC4.pdb (DDF)	2CD2.pdb (FOL)	3DFR.pdb (MTX)	
Thrombin			
1A2C.pdb (OAR)	1A4W.pdb (2EP)	1A61.pdb (LOM)	1AY6.pdb (HHO)
1BHX.pdb (R56)	1C5Z.pdb (BAM)	1D3D.pdb (BZT)	1D4P.pdb (BPP)
1D6W.pdb (AIM)	1EB1.pdb (ALC)	1FPC.pdb (ANS)	1G32.pdb (R11)
1GHW.pdb (BMZ)	1GI9.pdb (123)	1GJ4.pdb (132)	1GJ5.pdb (130)
1HDT.pdb (ALG)	1K1I.pdb (FD1)	1K21.pdb (IGN)	1K22.pdb (MEL)
8KME.pdb (TRG)			
TS			
1AXW.pdb (MTX)	1F28.pdb (F89)	1F4E.pdb (TPR)	1F4F.pdb (TP3)
1F4G.pdb (TP4)	1HVY.pdb (D16)	1JTQ.pdb (LY3)	1JTU.pdb (LYB)
1JUT.pdb (LYD)	1LCB.pdb (DHF)	1LCE.pdb (TMF)	1TDU.pdb (CB3)
1TSD.pdb (U18)	1TSL.pdb (A15)	2BBQ.pdb (PFG)	
PNP			
1A9P.pdb (9DI)	1A9R.pdb (HPA)	1A9S.pdb (NOS)	1A9T.pdb (R1P)
1B8N.pdb (IMG)	1B8O.pdb (IMH)	1C3X.pdb (8IG)	1K9S.pdb (FM1/2)
1FXU.pdb (GU7)	1I80.pdb (IMR)	1ULB.pdb (GUN)	1CG6.pdb (MTA)
AChE			
1ACJ.pdb (THA)	1E3Q.pdb (EBW)	1E66.pdb (HUX)	1GPN.pdb (HUB)
1GQR.pdb (SAF)	1HBJ.pdb (FBQ)	1QON.pdb (I40)	2ACK.pdb (EDR)

**Table S2.** Experimentally tested L99A ligands.

L99A LIGANDS			
1		16	
2		17	
3		18	
4		19	
5		20	
6		21	
7		22	
8		23	
9		24	
10		25	
11		26	
12		27	
13		28	
14		29	
15		30	
31		31	
32		32	
33		33	
34		34	
35		35	
36		36	
37		37	
38		38	
39		39	
40		40	
41		41	
42		42	
43		43	
44		44	
45		45	
46		46	
47		47	
48		48	
49		49	
50		50	
51		51	
52		52	
53		53	
54		54	
55		55	

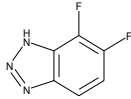
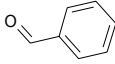



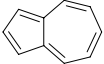
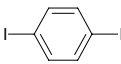



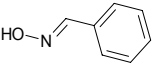
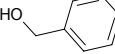
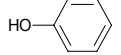
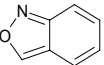


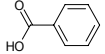


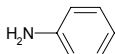
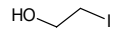
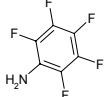
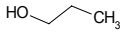
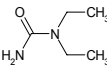
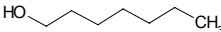
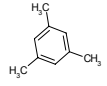
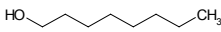
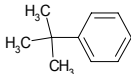


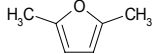
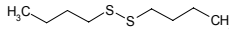
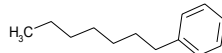








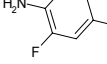



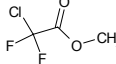
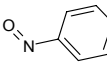

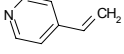
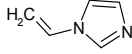
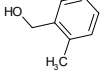
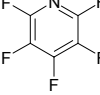
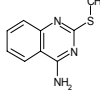
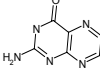


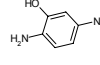




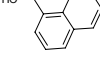

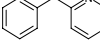


**Table S3.** Experimentally tested L99A/M102Q ligands.<sup>a</sup>

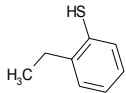
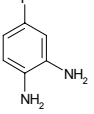
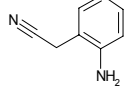
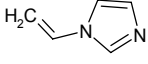
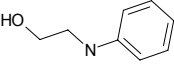
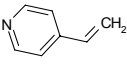
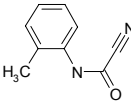
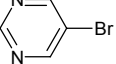
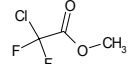

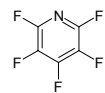
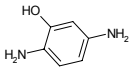
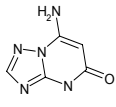
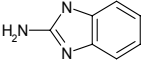
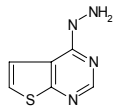
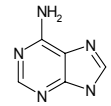
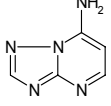
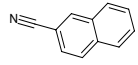
L99A/M102Q LIGANDS			
1		7	
2		8	
3		9	
4		10	
5		11	
6		12	
		13	
		14	
		15	
		16	
		17	
		18	
		19	
		20	
		21	
		22	
		23	

<sup>a</sup>L99A ligands (Table S2) are also ligands for L99A/M102Q

**Table S4.** Experimentally tested L99A decoys.

L99A DECOYS			
1		17	
2		18	
3		19	
4		20	
5		21	
6		22	
7		23	
8		24	
9		25	
10		26	
11		27	
12		28	
13		29	
14		30	
15		31	
16		32	
		33	
		34	
		35	
		36	
		37	
		38	
		39	
		40	
		41	
		42	
		43	
		44	
		45	
		46	
		47	
		48	
		49	
		50	
		51	
		52	
		53	
		54	
		55	
		56	
		57	
		58	
		59	
		60	
		61	
		62	
		63	
		64	

**Table S5.** Experimentally tested L99A/M102Q decoys.

L99A/M102Q DECOYS			
1		6	
2		7	
3		8	
4		9	
5		10	
		11	
		12	
		13	
		14	
		15	
		16	
		17	
		18	

**Table S6.** Experimentally tested AmpC ligands and decoys.

AMPIC LIGANDS			
1		3	
2		4	
5		6	
7			
AMPIC DECOYS			
1		22	
2		23	
3		24	
4		25	
5		26	
6		27	
7		28	
8		29?s	
9		30	
10		31	
11		32	
12		33	
13		34	
14		35	
15		36f	
43		44	
45		46	
47		48	
49		50	
51		52	
53		54	
55		56	
57		58	
64		65	
66		67	
68		69	
70		71	
72		73	
74		75	
76		77	
78			

**Table S6 (continued).** Experimentally tested AmpC ligands and decoys.

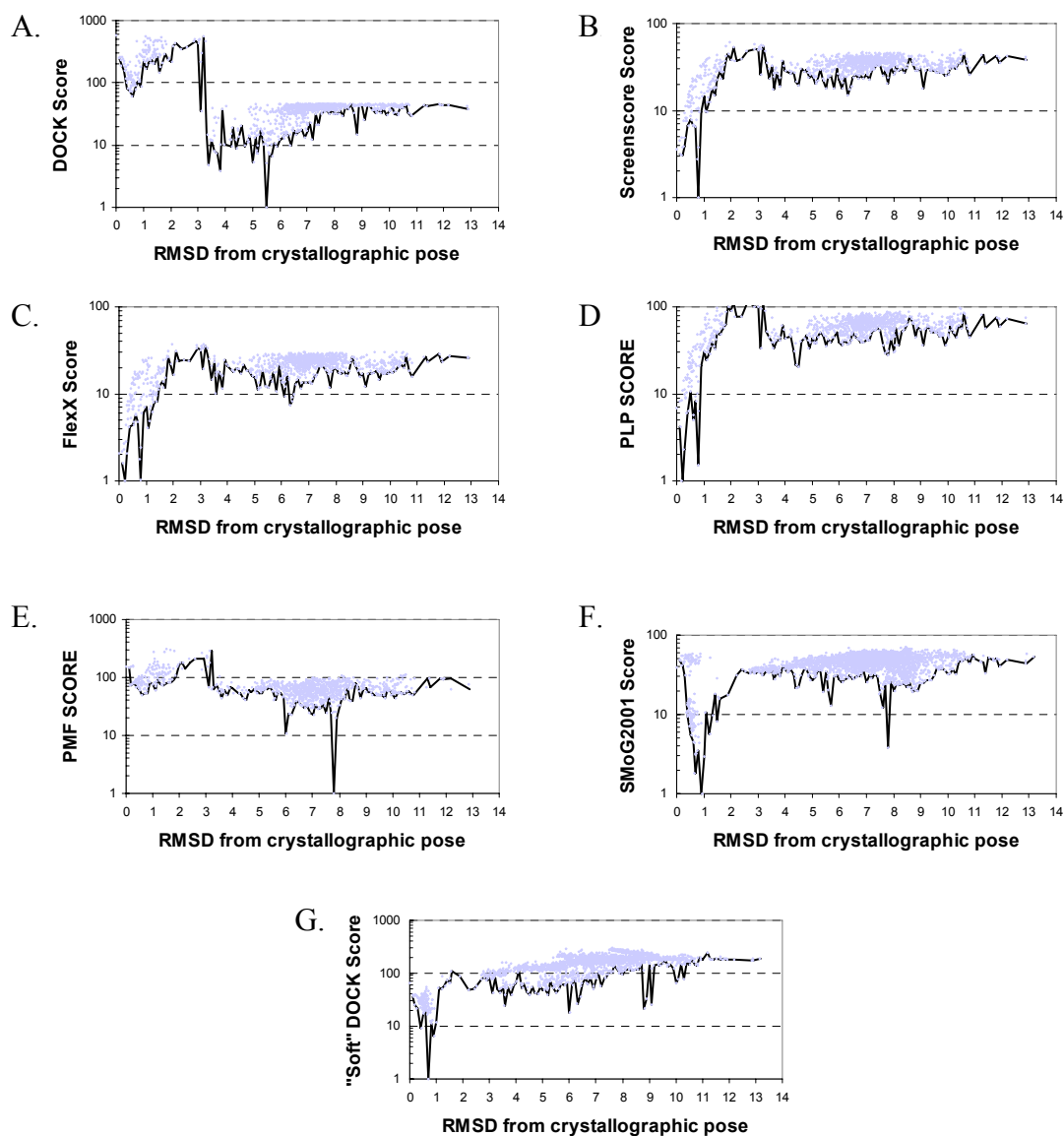
16		37		58		79	
17		38		59		80	
18d		39		60		81	
19		40		61		82	
20		41		62		83	
21		42		63		84	

**Table S7.** Crystal data for catechol in complex with L99A/M102Q

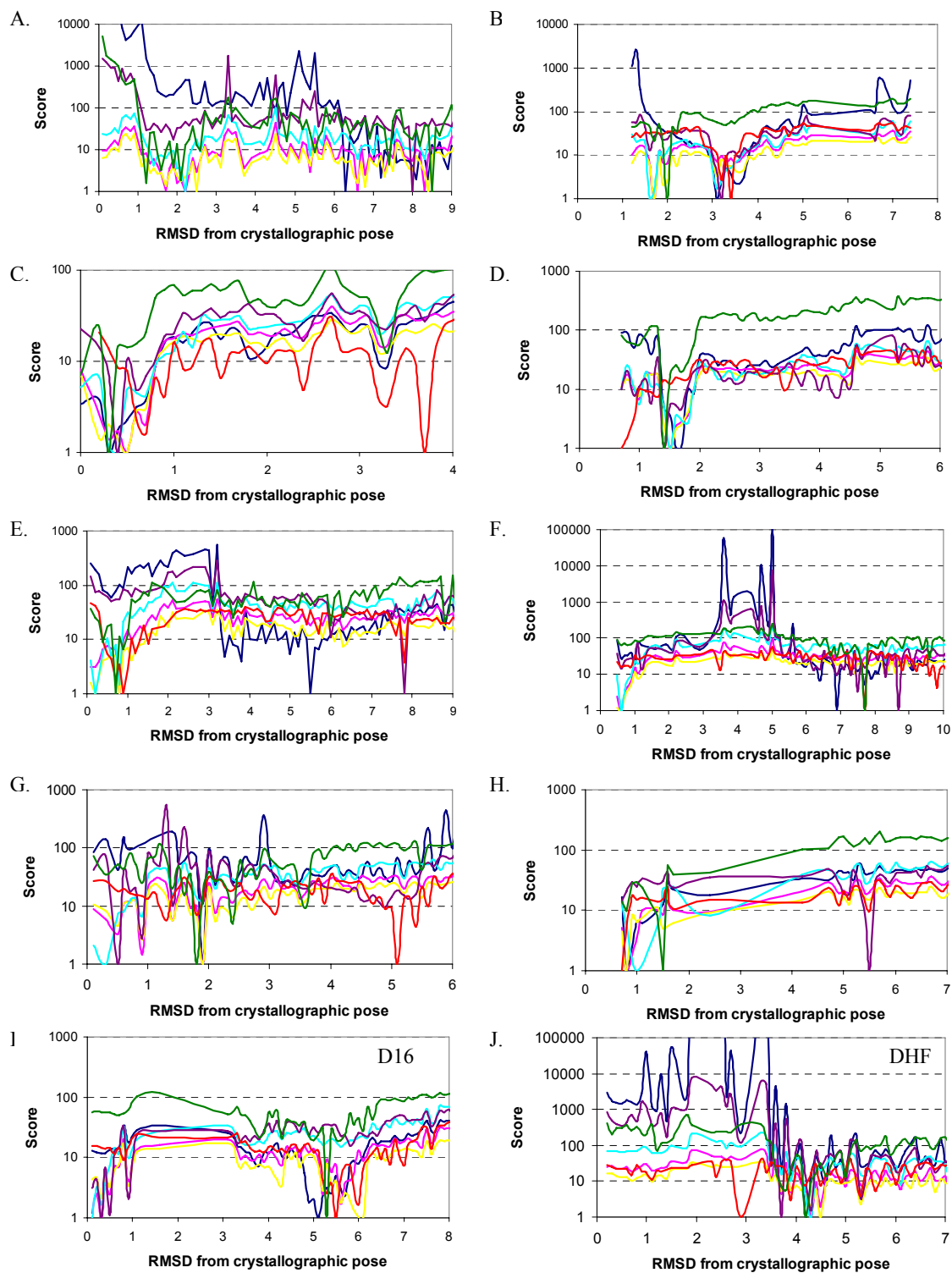
Space group	P3 <sub>2</sub> 21
Cell constants	
a (Å)	60.17
c (Å)	96.56
Resolution (Å)	40 – 1.553
Total No. of reflections	413,552
No. of unique reflections	29,343
Completeness of all data (% , outer shell)	97.7 (97.6)
R <sub>symm</sub> for all data (% , outer shell) <sup>a</sup>	4.0 (56.5)
R <sub>free</sub> (%) <sup>b</sup>	21.9
R-factor (%) <sup>b</sup>	18.8
No. of protein residues	162
No. of water molecules	227
r.m.s. dev., angle (deg.)	1.3
r.m.s. dev., bond (Å)	0.009
Average <i>B</i> -factors, protein atoms (Å <sup>2</sup> )	17.7
Average <i>B</i> -factors, water atoms (Å <sup>2</sup> )	30.5
Average <i>B</i> -factors, ligand atoms (Å <sup>2</sup> )	18.4

<sup>a</sup>R<sub>symm</sub> =  $\sum |I - \langle I \rangle| / \sum I$ , where *I* is the observed intensity and  $\langle I \rangle$  is the average intensity for multiple measurements.

<sup>b</sup>The R<sub>free</sub> was calculated from a random selection of reflections constituting 5 % of the data; the R-factor was calculated with the remaining intensities.

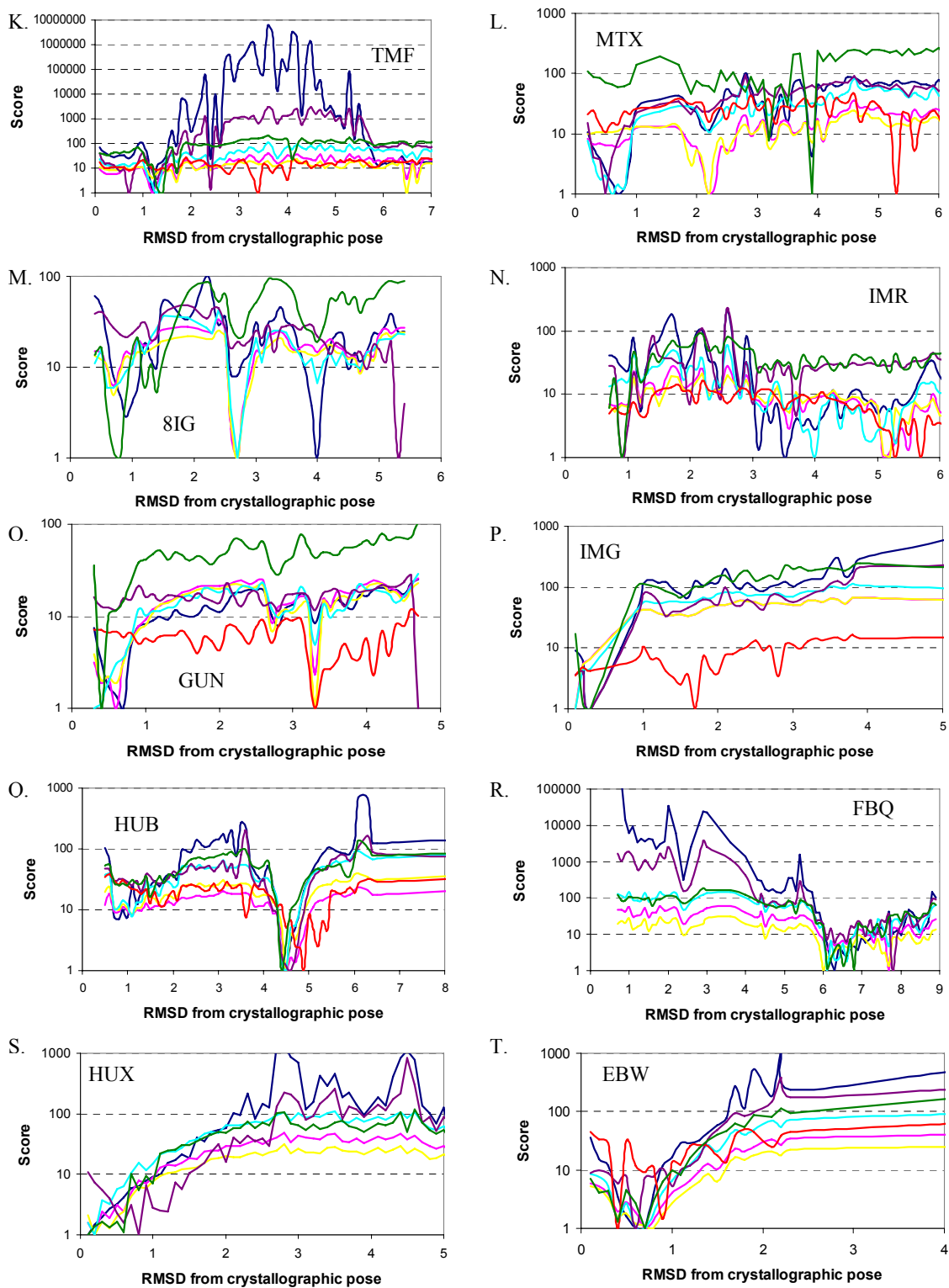


**Figure S1.** Score vs. RMSD from native pose for ligand ALG in thrombin using scoring functions from A. DOCK, B. ScreenScore, C. FlexX, D. PLP, E. PMF, F. SMOG, and G. DOCK score with an 8-6 van der Waals potential and 4 fold increase in electrostatic interaction energy. Curves bounding the low energy poses are shown.

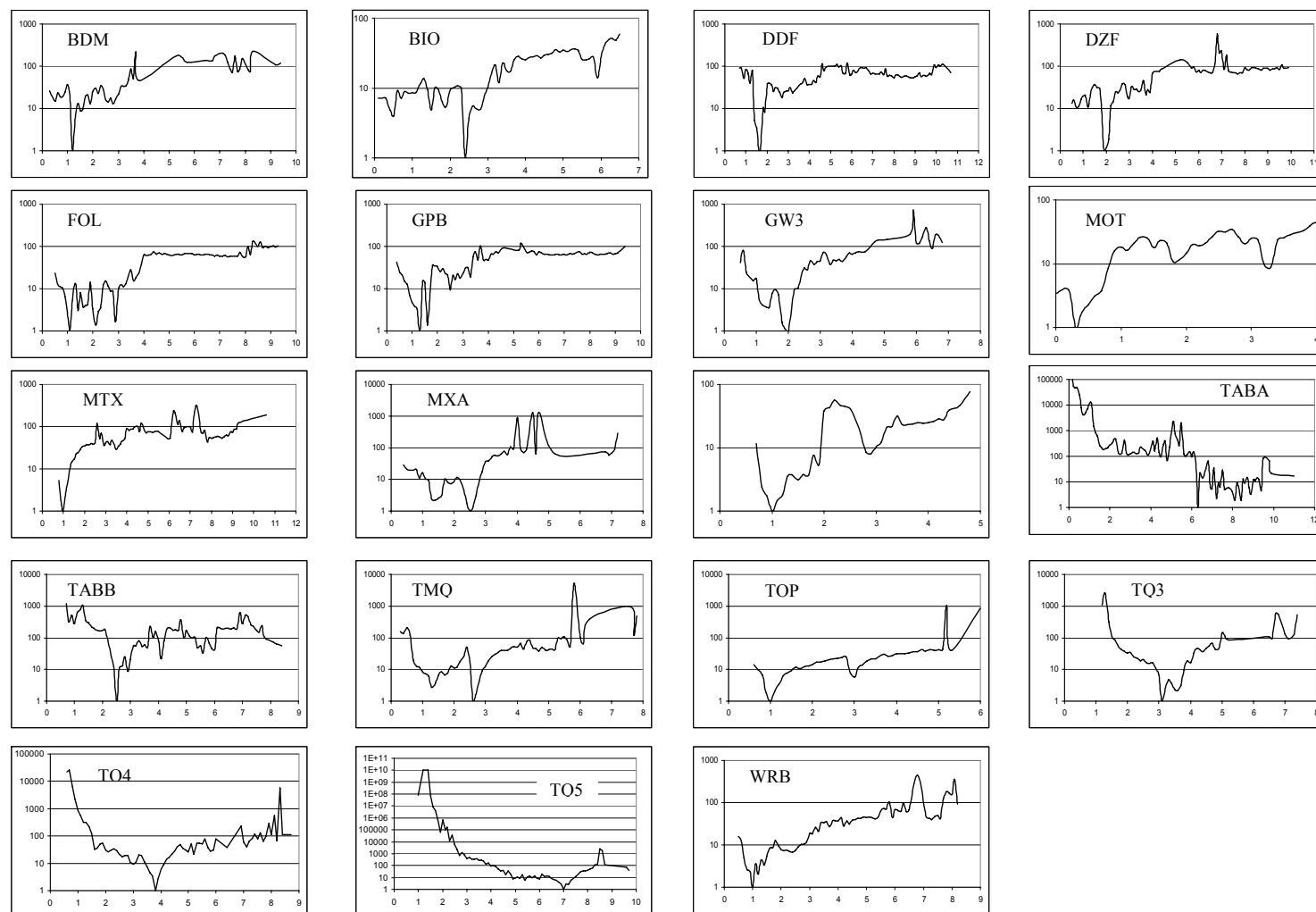


**Figure S2.** Score vs. RMSD bounding curves for native-like and geometric decoys from Chapter 1. Table 1.

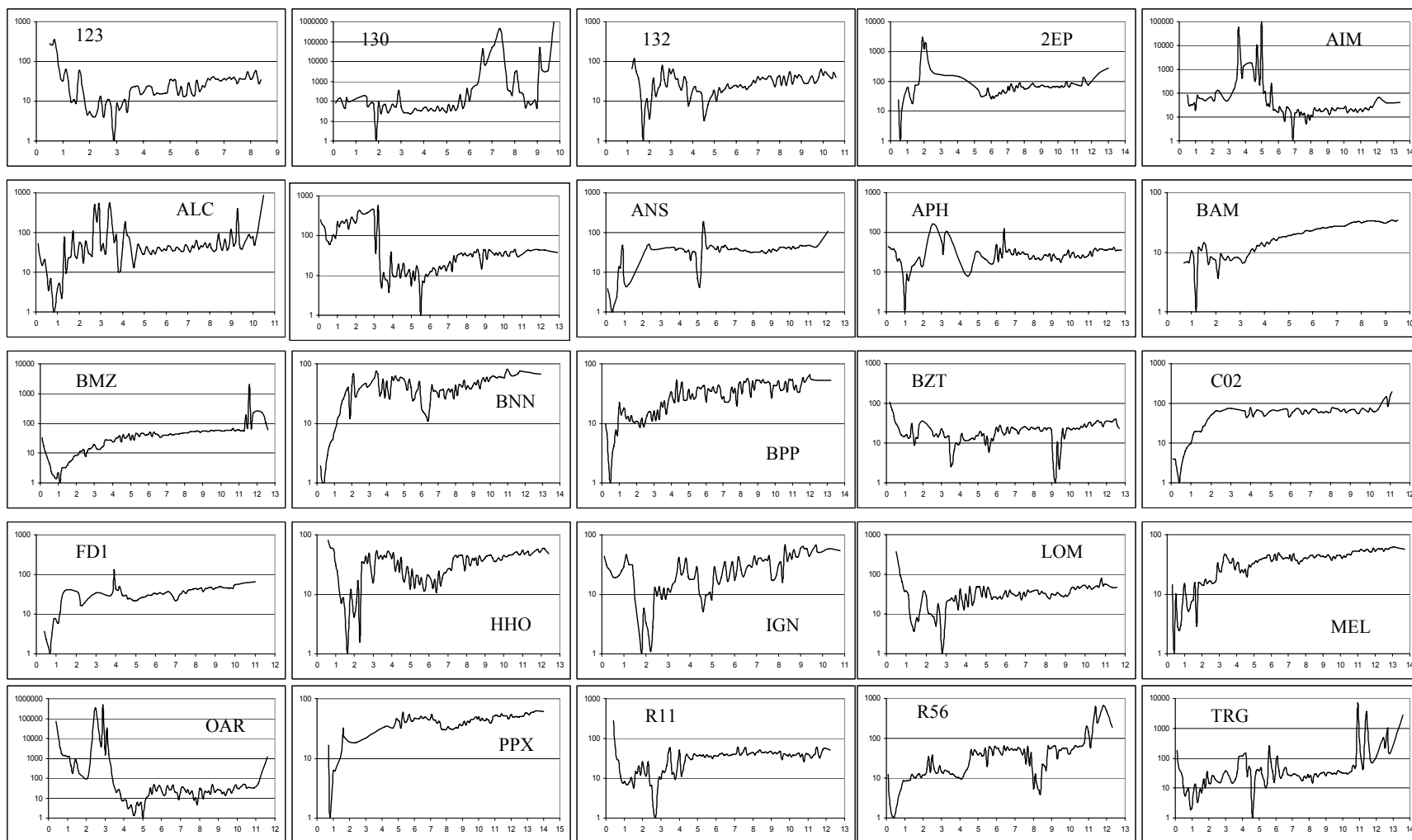




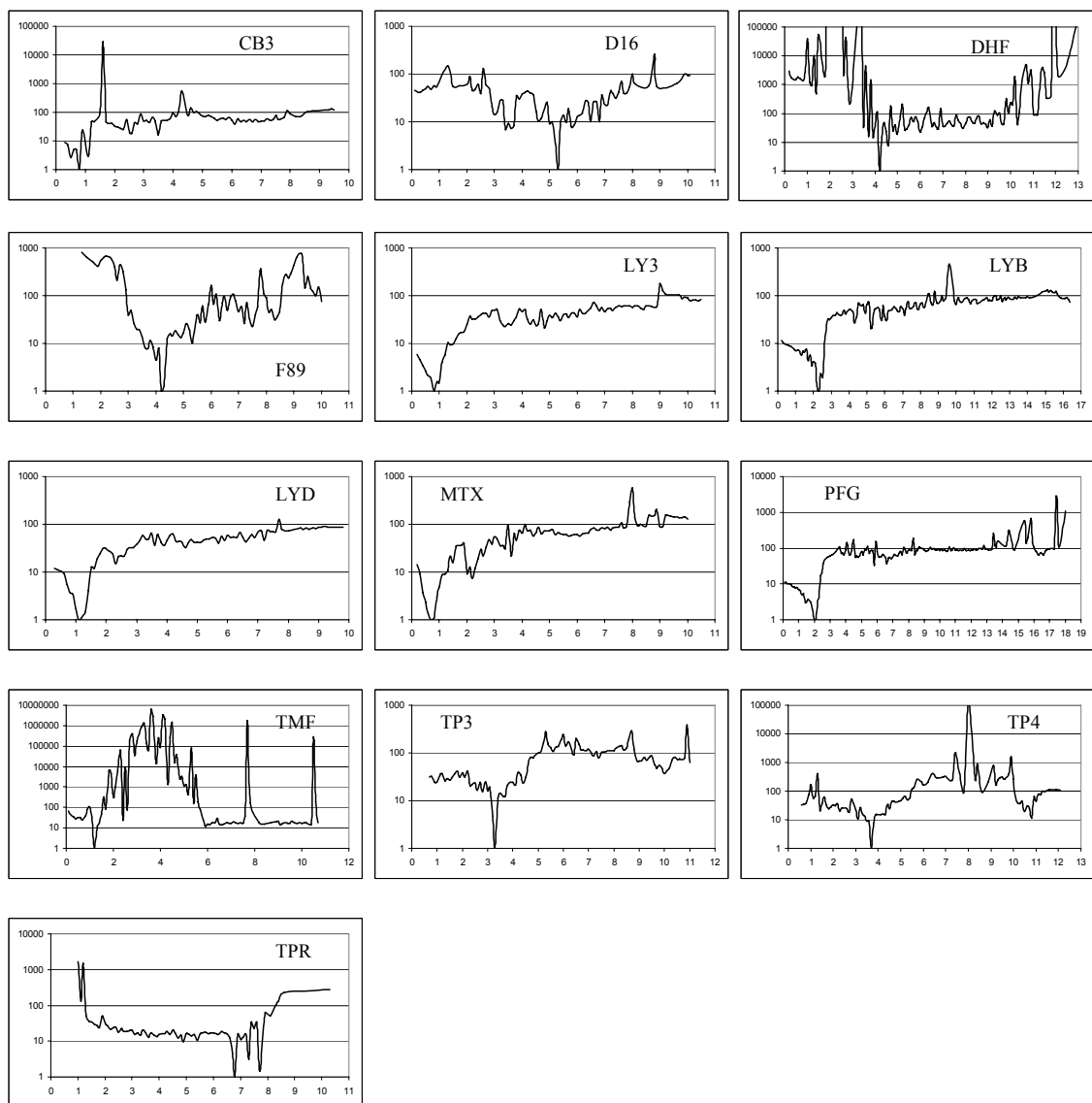
**Figure S2 (continued).** Score vs. RMSD bounding curves for native-like and geometric decoys from Chapter 1. Table 1.



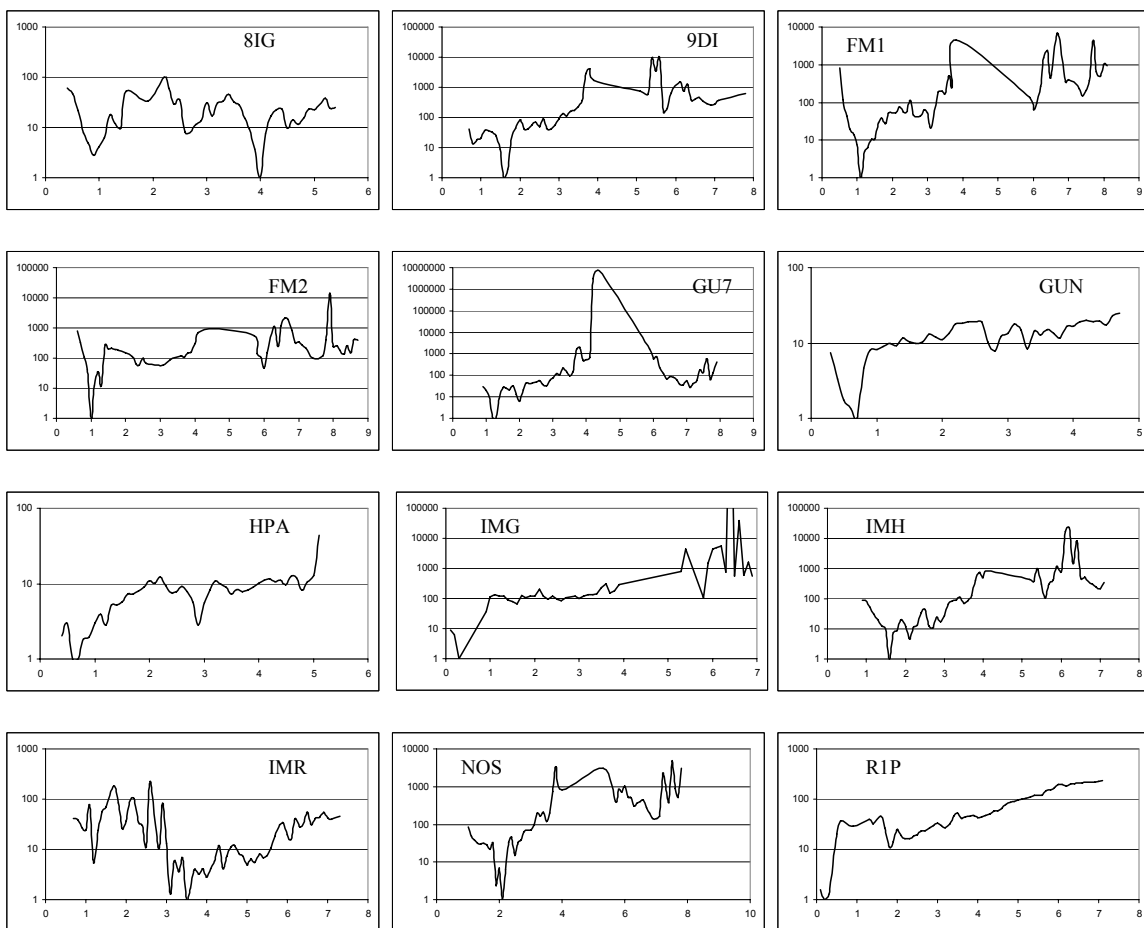
**Figure S3.** Minimum bounding curves of DOCK score vs. RMSD from crystallographic pose for DHFR ligands.



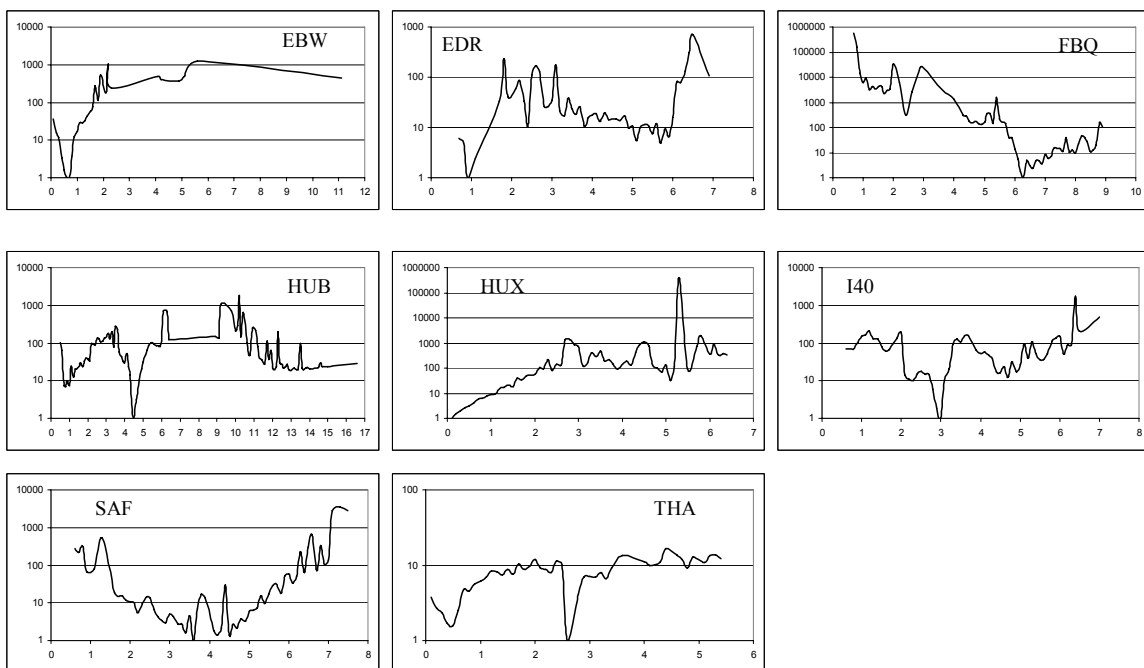
**Figure S4.** Minimum bounding curves of DOCK score vs. RMSD from crystallographic pose for thrombin ligands.



**Figure S5.** Minimum bounding curves of DOCK score vs. RMSD from crystallographic pose for TS ligands.



**Figure S6.** Minimum bounding curves of DOCK score vs. RMSD from crystallographic pose for PNP ligands.



**Figure S7.** Minimum bounding curves of DOCK score vs. RMSD from crystallographic pose for AChE ligands.

## **Appendix B.**

### **Supporting Material**

#### **Rescoring Docking Hit Lists for Model Cavity Sites: Predictions and Experimental Testing**

Alan P. Graves, Devleena M. Shivakumar, Sarah E. Boyce, Matthew P. Jacobson, David  
A. Case, and Brian K. Shoichet

#### **Contents:**

1. PLOP side chain rotamer search and minimization algorithm.
2. AMBERDOCK parameters and optimization.
3. Tables S1-S9.
4. Figures S1 and S2.

**PLOP side chain rotamer search and minimization algorithm.** The side chain rotamer search algorithm proceeds by optimizing one side chain at a time (keeping the others fixed), and keeps looping through the residues until they stop changing conformation. During the side chain rotamer search, the default resolution of  $10^\circ$  for the rotamer library was used, and the initial side chain conformations were utilized by specifying the parameter “randomize no”. The initial (ofac\_init) and minimal (ofac\_min) overlap factors, which define steric clashes, were set to 0.75 and 0.5, respectively. The calculation of the complex energy ( $E_{complex}$ ) involved an initial short minimization of the ligand with a maximum number of steps (mxitn) set to 15 followed by the sidechain rotamer search, minimization of residues with a maximum number of steps (mxitn) set to 30, and finally minimization of ligand to convergence. Minimization and calculation of the free receptor energy ( $E_{receptor}$ ) began with the initial side chain rotamer search followed by the minimization of residues with a maximum number of steps (mxitn) set to 30.

**AMBERDOCK parameters and optimization.** The traditional AMBER force field has been parameterized to work with biological macromolecules, and has limited parameters for organic molecules. The general AMBER force field (GAFF), which has been specifically designed to cover most pharmaceutical molecules, is compatible with the existing AMBER force fields in such a way that the two can be mixed during a simulation. It uses a simple harmonic energy function as the additive AMBER force fields, but the atom types used in GAFF are much more general such that they cover most of the organic chemical space. The current implementation of the GAFF force field consists of 33 basic atom types and 22 special atom types covering almost all of the organic chemical space that is made up of C, N, O, S, P, H, F, Cl, Br and I atoms. The input ligand files for AMBERDOCK can be generated automatically using the perl script (prepare\_amber.pl) provided with the DOCK6.1 program. This perl script calls for



programs such as *antechamber* to calculate the charges for the ligands and *tleap* to assign the parameter set for protein and ligand atoms. The newer version of the DOCK6 program bypasses this external perl step, and the input files are generated internally when the AMBER score routine is called.

Binding free energy calculations with AMBERDOCK follows the flowchart in Figure S1. In Step 1, the input files (pdb, inpcrd, prmtop) in AMBER format are read. In Step 2, the frozen atoms are defined based on the choice of the user. The user can specify the atoms that are allowed to relax in the DOCK6 input file (dock.in). A brief minimization is carried out using the GB implicit solvent model in Step 2. In Step 3, the molecular dynamics (MD) simulation is carried out on the minimized system. In step 4, a brief minimization is carried out to minimize the structure generated from the previous MD step. Steps 2-4 are performed with the frozen atom approximation in GB implicit solvent model. When frozen atoms are specified, the NAB program calculates the energy for only those atoms that are allowed to move. This significantly speeds up the calculations and uses less memory. In Step 5, a single point energy is calculated on the structure obtained from Step 4. In this step, the energy of the full system is calculated without any frozen atoms and non-bonded cutoffs. Also, the surface area term is added to this final step to get a more accurate energy term. The Steps 1-5 were repeated for complex ( $E_{complex}$ ), ligand ( $E_{ligand}$ ) and receptor ( $E_{receptor}$ ). To expedite the scoring process, we calculated the energy of the receptor for the first ligand and used this energy as a constant term during the energy evaluations for the rest of the ligands in the database. It is suggested to repeat these steps with various combinations of molecular mechanics (MM) options to optimize the variables.

Several scoring protocols were tested on the 58 known ligands and 17 known decoys for L99A/M102Q to find the optimal set of conditions for AMBERDOCK rescoring. The initial attempt was to use AMBERDOCK with only minimization and no molecular dynamics simulation. For example, *score1* (Figure S2b) involved scoring the

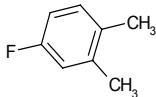
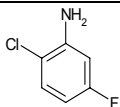
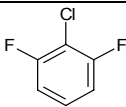
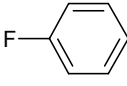
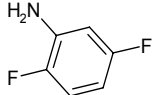
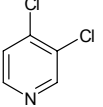
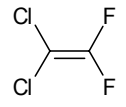
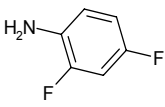
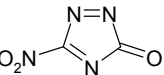
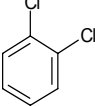
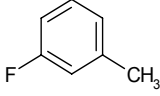
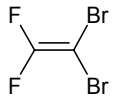
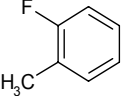
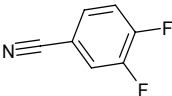
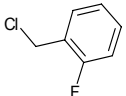
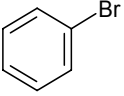
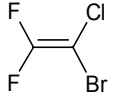
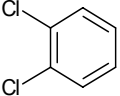
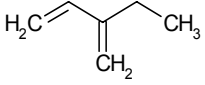
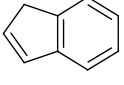
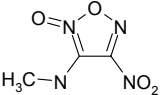
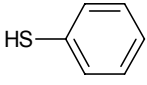
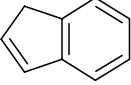
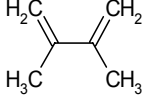
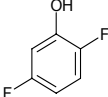
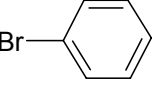
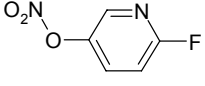
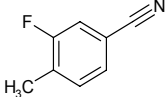
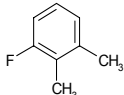
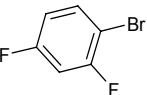
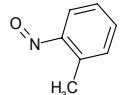
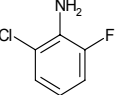

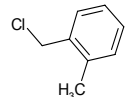
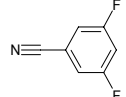
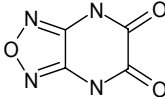
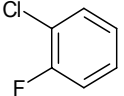
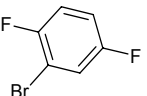
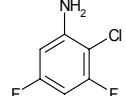
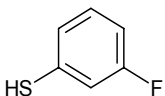
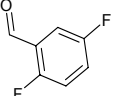
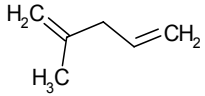
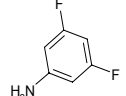
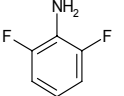
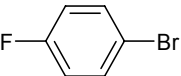
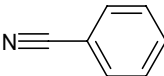
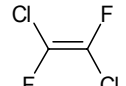
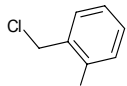
test set with Gasteiger-Marselli PEOE charges for the ligands and AMBER parm 94 parameter set for protein atoms; the Hawkins-Cramer-Truhlar pairwise GB model (equivalent to  $igb=1$  in AMBER); the LCPO method to calculate the surface area term; and minimization of nine binding site residues (78, 84, 88, 91, 102, 111, 118, 121, and 153). The results clearly show that the minimization alone is not sufficient to separate out the ligands from decoys with several known decoys scoring better than known ligands. Infact, the results from DOCK (Figure S2a) is adversely affected by running *score1*. We also tried rescoring the same test set with a series of AMBERDOCK scores with different combinations of the molecular mechanics input parameters, using large or small flexible receptor regions, GB models ( $igb=1, 2, 5$ ) with or without the surface area term, number of minimization steps, etc. We found that all of them predicted several decoys among the top 20 scored ligands with only minimization of the system, and no molecular dynamics (MD) simulations. As a logical next step, we introduced MD simulations along with minimization in the system. The initial receptor-ligand configuration obtained from docking was subjected to a few steps of minimization followed by a few picoseconds of MD simulations, and then a final minimization run to obtain the total energy. Selection of charge models is also very important. The AMBERDOCK scores obtained when using the AM1-BCC charge model provided better results compared to the Gasteiger-Marselli PEOE charge model. Using *score24G* with Gasteiger-Marselli PEOE charges with MD and minimization, 6 decoys scored as well as the top 20 ligands (Figure S2c). Using *score24* with AM1-BCC charges with MD and minimization, only 1 decoy scored as well as the top 20 scoring ligands (Figure S2d). This scoring protocol produced good results for most of our test sets of known ligands and decoys compared to the other scoring protocols that we tested. The specifications of *score24* are: GB model of 5 (corresponding to  $igb=5$  in the AMBER program); surface area term calculated using the LCPO model; a non-bonded cutoff of 18 Å; 100 steps of conjugate gradient minimization; and 3000 steps of MD simulation with a 1 fs time step

at a temperature of 300K, followed by 100 steps of minimization. During the minimization and MD, only the ligand and the protein residues within 5 Å of the ligand were allowed to move.

**Table S1.** Top 100 hits predicted by DOCK3.5.54 for L99A.

#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
1		14		27		40	
2		15		28		41	
3		16		29		42	
4		17		30		43	
5		18		31		44	
6		19		32		45	
7		20		33		46	
8		21		34		47	
9		22		35		48	
10		23		36		49	
11		24		37		50	
12		25		38		51	
13		26		39		52	

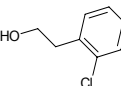

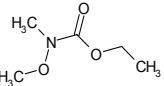
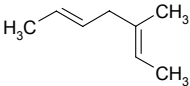
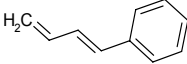
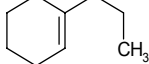
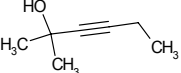
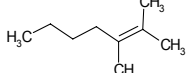
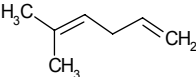
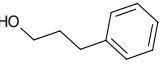
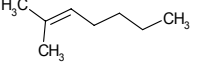
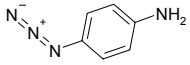
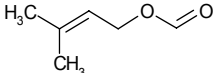
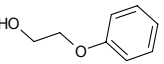
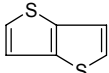
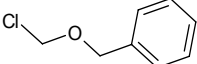
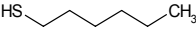
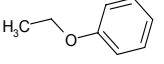
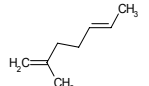
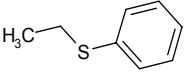

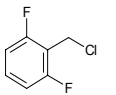
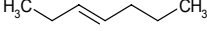
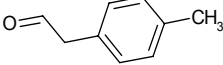

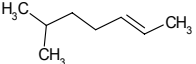
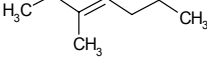
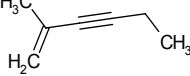
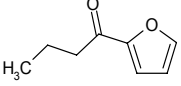
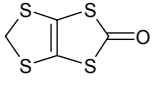
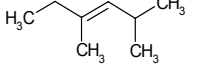
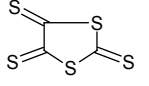
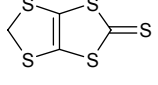
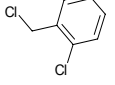
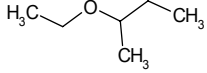
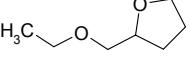
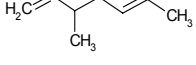
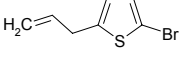
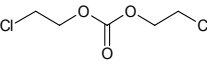
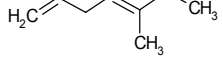

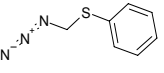
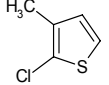
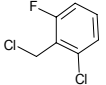
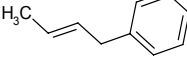
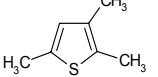
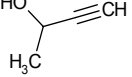
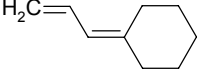
**Table S1 (continued).** Top 100 hits predicted by DOCK3.5.54 for L99A.

#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
53		65		77		89	
54		66		78		90	
55		67		79		91	
56		68		80		92	
57		69		81		93	
58		70		82		94	
59		71		83		95	
60		72		84		96	
61		73		85		97	
62		74		86		98	
63		75		87		99	
64		76		88		100	

**Table S2.** Top 100 hits predicted by PLOP for L99A.

#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
1		14		27		40	
2		15		28		41	
3		16		29		42	
4		17		30		43	
5		18		31		44	
6		19		32		45	
7		20		33		46	
8		21		34		47	
9		22		35		48	
10		23		36		49	
11		24		37		50	
12		25		38		51	
13		26		39		52	

**Table S2 (continued).** Top 100 hits predicted by PLOP for L99A.

#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
53		65		77		89	
54		66		78		90	
55		67		79		91	
56		68		80		92	
57		69		81		93	
58		70		82		94	
59		71		83		95	
60		72		84		96	
61		73		85		97	
62		74		86		98	
63		75		87		99	
64		76		88		100	

**Table S3.** Top 100 hits predicted by AMBERDOCK for L99A.

#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
1		14		27		40	
2		15		28		41	
3		16		29		42	
4		17		30		43	
5		18		31		44	
6		19		32		45	
7		20		33		46	
8		21		34		47	
9		22		35		48	
10		23		36		49	
11		24		37		50	
12		25		38		51	
13		26		39		52	



**Table S3 (continued).** Top 100 hits predicted by AMBERDOCK for L99A.

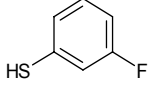
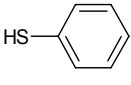
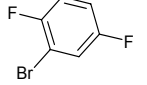
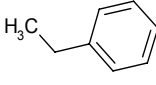
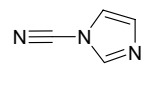
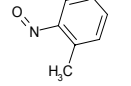

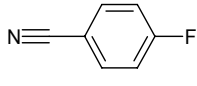
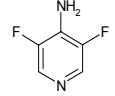
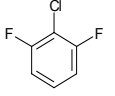
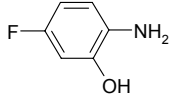
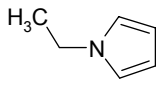
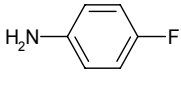
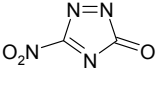

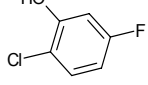
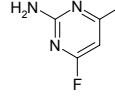
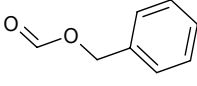
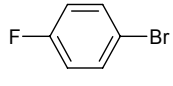
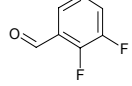
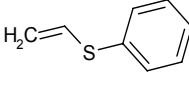
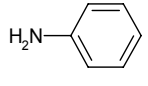
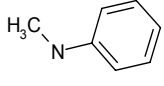
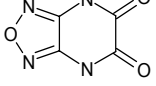
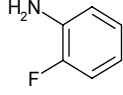
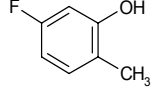
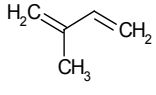
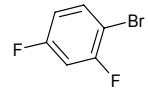
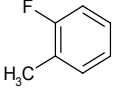
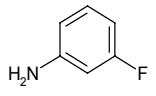
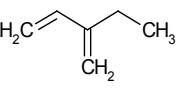
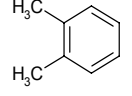
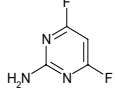
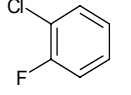
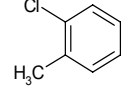
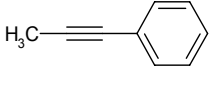
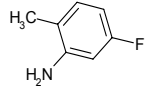
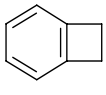
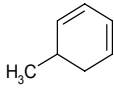
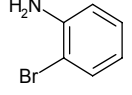
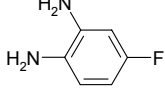
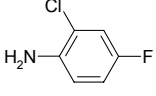
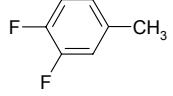
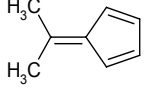
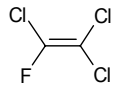
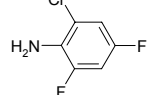
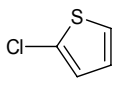
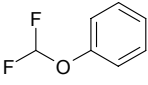
#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
53		65		77		89	
54		66		78		90	
55		67		79		91	
56		68		80		92	
57		69		81		93	
58		70		82		94	
59		71		83		95	
60		72		84		96	
61		73		85		97	
62		74		86		98	
63		75		87		99	
64		76		88		100	

**Table S4.** Top 100 hits predicted by DOCK3.5.45 for L99A/M102Q.

#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
1		14		27		40	
2		15		28		41	
3		16		29		42	
4		17		30		43	
5		18		31		44	
6		19		32		45	
7		20		33		46	
8		21		34		47	
9		22		35		48	
10		23		36		49	
11		24		37		50	
12		25		38		51	

**Table S4 (continued).** Top 100 hits predicted by DOCK3.5.54 for

L99A/M102Q.

#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
53		65		77		89	
54		66		78		90	
55		67		79		91	
56		68		80		92	
57		69		81		93	
58		70		82		94	
59		71		83		95	
60		72		84		96	
61		73		85		97	
62		74		86		98	
63		75		87		99	
64		76		88		100	

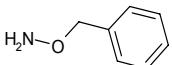
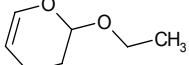
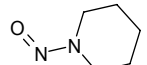
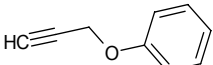
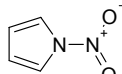
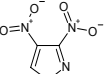
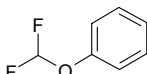
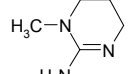
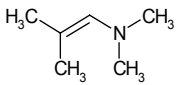
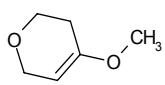
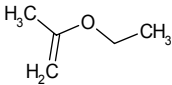
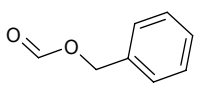
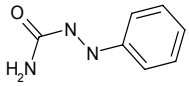
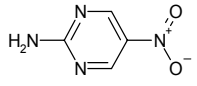
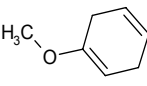
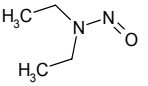
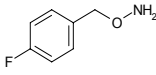
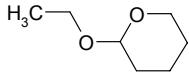
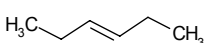
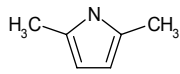
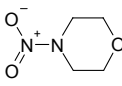
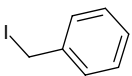
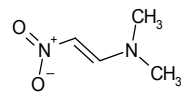
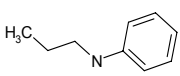
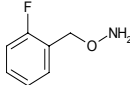
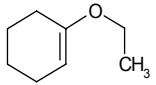
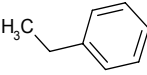
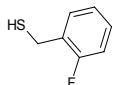
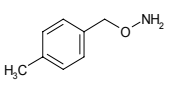
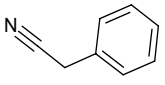
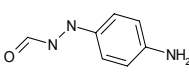
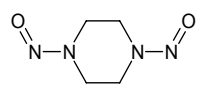
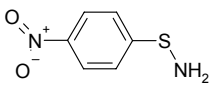
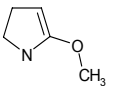
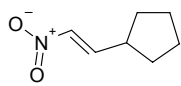
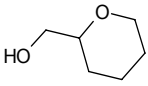
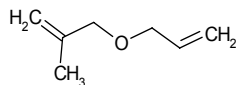
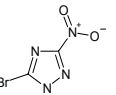
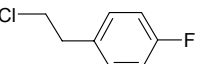
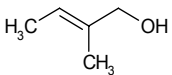
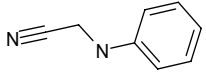
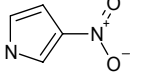
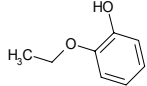
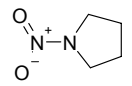

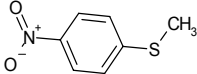
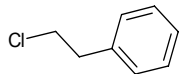
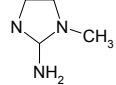
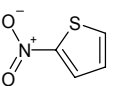
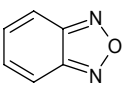
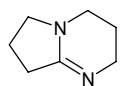

**Table S5.** Top 100 hits predicted by PLOP for L99A/M102Q.

#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
1		14		27		40	
2		15		28		41	
3		16		29		42	
4		17		30		43	
5		18		31		44	
6		19		32		45	
7		20		33		46	
8		21		34		47	
9		22		35		48	
10		23		36		49	
11		24		37		50	
12		25		38		51	
13		26		39		52	

**Table S5 (continued).** Top 100 hits predicted by PLOP for L99A/M102Q.

#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
53		65		77		89	
54		66		78		90	
55		67		79		91	
56		68		80		92	
57		69		81		93	
58		70		82		94	
59		71		83		95	
60		72		84		96	
61		73		85		97	
62		74		86		98	
63		75		87		99	
64		76		88		100	

**Table S6.** Top 100 hits predicted by AMBERDOCK for L99A/M102Q.

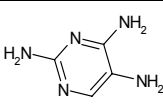
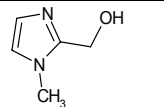
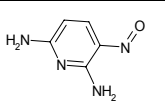
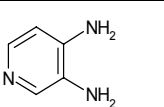
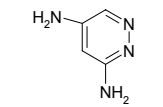
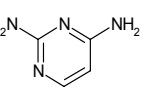
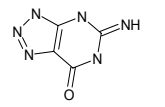
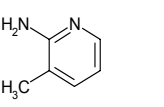
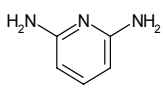
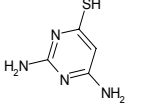
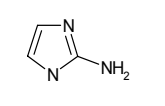
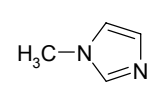
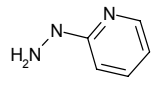
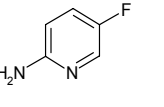
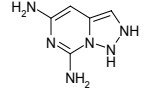
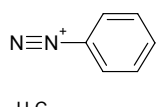
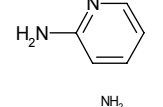
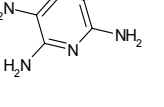
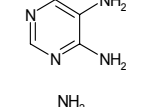
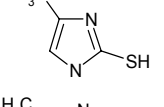
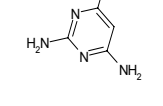
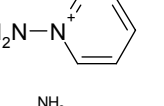
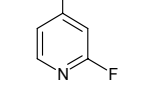
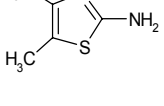
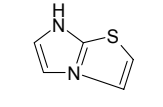
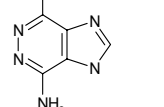
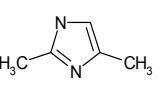
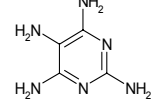
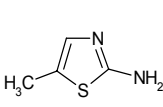
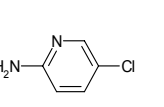
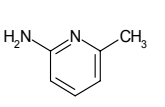
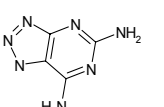
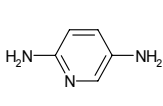
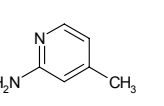
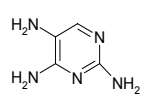
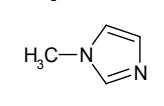
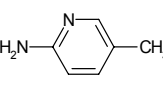
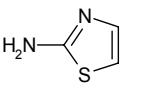
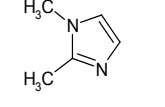
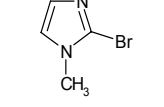
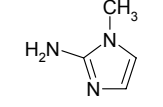
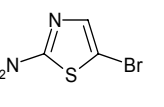
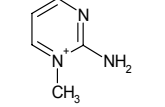
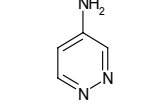
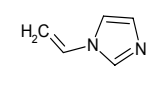
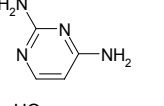
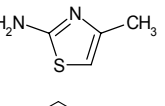
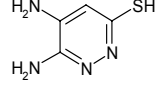
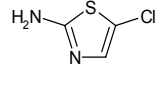
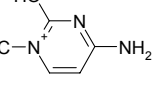
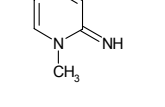
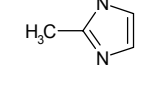
#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
1		14		27		40	
2		15		28		41	
3		16		29		42	
4		17		30		43	
5		18		31		44	
6		19		32		45	
7		20		33		46	
8		21		34		47	
9		22		35		48	
10		23		36		49	
11		24		37		50	
12		25		38		51	
13		26		39		52	

**Table S6 (continued).** Top 100 hits predicted by AMBERDOCK for

L99A/M102Q.

#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
53		65		77		89	
54		66		78		90	
55		67		79		91	
56		68		80		92	
57		69		81		93	
58		70		82		94	
59		71		83		95	
60		72		84		96	
61		73		85		97	
62		74		86		98	
63		75		87		99	
64		76		88		100	

**Table S7.** Top 100 hits predicted by DOCK3.5.54 for CCP.

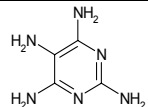
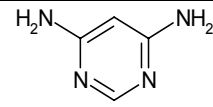
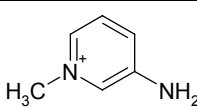
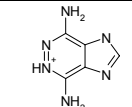
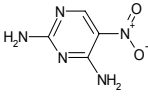
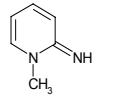
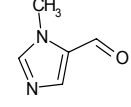
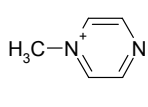
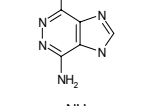
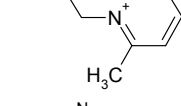
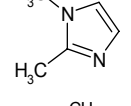
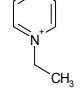
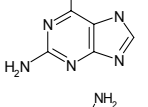
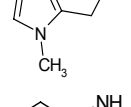
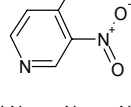
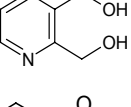
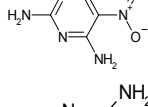
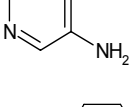
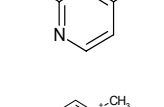
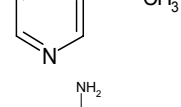
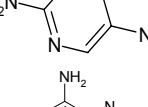
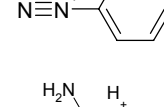
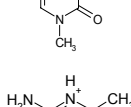
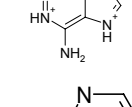
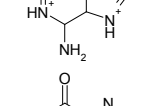
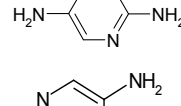
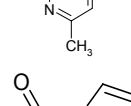
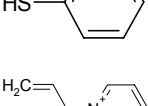
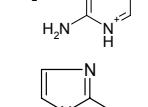
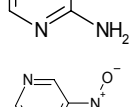
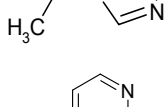
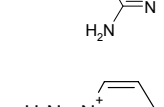
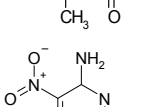
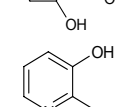
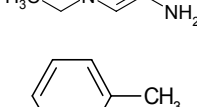
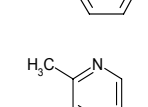
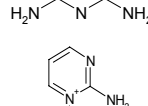
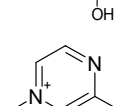
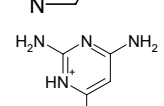
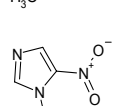
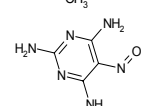
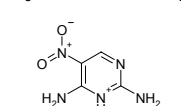
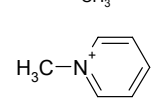
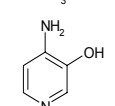
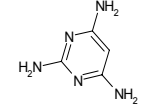
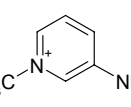
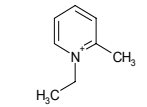
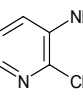




#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
1		14		27		40	
2		15		28		41	
3		16		29		42	
4		17		30		43	
5		18		31		44	
6		19		32		45	
7		20		33		46	
8		21		34		47	
9		22		35		48	
10		23		36		49	
11		24		37		50	
12		25		38		51	
13		26		39		52	



**Table S7 (continued).** Top 100 hits predicted by DOCK3.5.54 for CCP.

#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
53		65		77		89	
54		66		78		90	
55		67		79		91	
56		68		80		92	
57		69		81		93	
58		70		82		94	
59		71		83		95	
60		72		84		96	
61		73		85		97	
62		74		86		98	
63		75		87		99	
64		76		88		100	

**Table S8.** Top 100 hits predicted by PLOP for CCP.

#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
1		14		27		40	
2		15		28		41	
3		16		29		42	
4		17		30		43	
5		18		31		44	
6		19		32		45	
7		20		33		46	
8		21		34		47	
9		22		35		48	
10		23		36		49	
11		24		37		50	
12		25		38		51	
13		26		39		52	

**Table S8 (continued).** Top 100 hits predicted by PLOP for CCP.

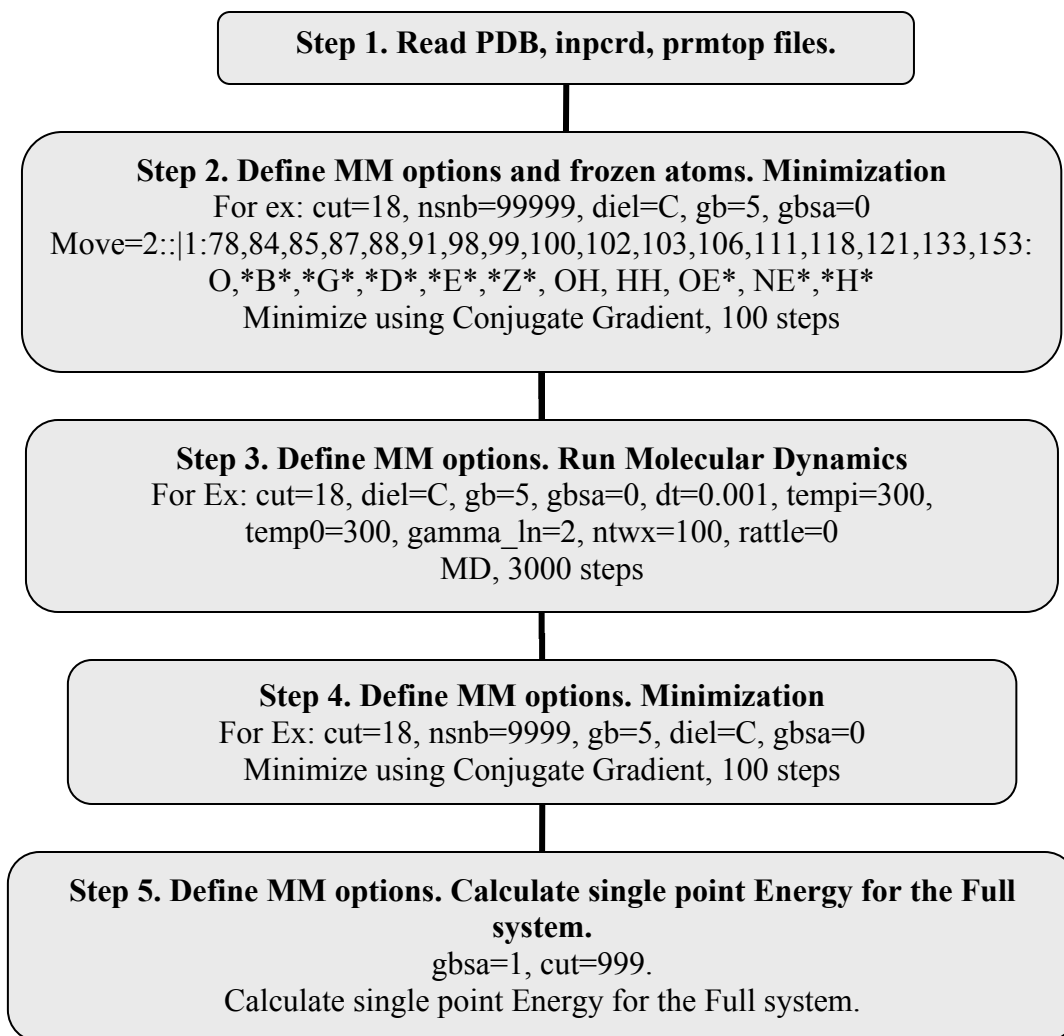
#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
53		65		77		89	
54		66		78		90	
55		67		79		91	
56		68		80		92	
57		69		81		93	
58		70		82		94	
59		71		83		95	
60		72		84		96	
61		73		85		97	
62		74		86		98	
63		75		87		99	
64		76		88		100	

**Table S9.** Top 100 hits predicted by AMBERDOCK for CCP.

#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
1		14		27		40	
2		15		28		41	
3		16		29		42	
4		17		30		43	
5		18		31		44	
6		19		32		45	
7		20		33		46	
8		21		34		47	
9		22		35		48	
10		23		36		49	
11		24		37		50	
12		25		38		51	
13		26		39		52	

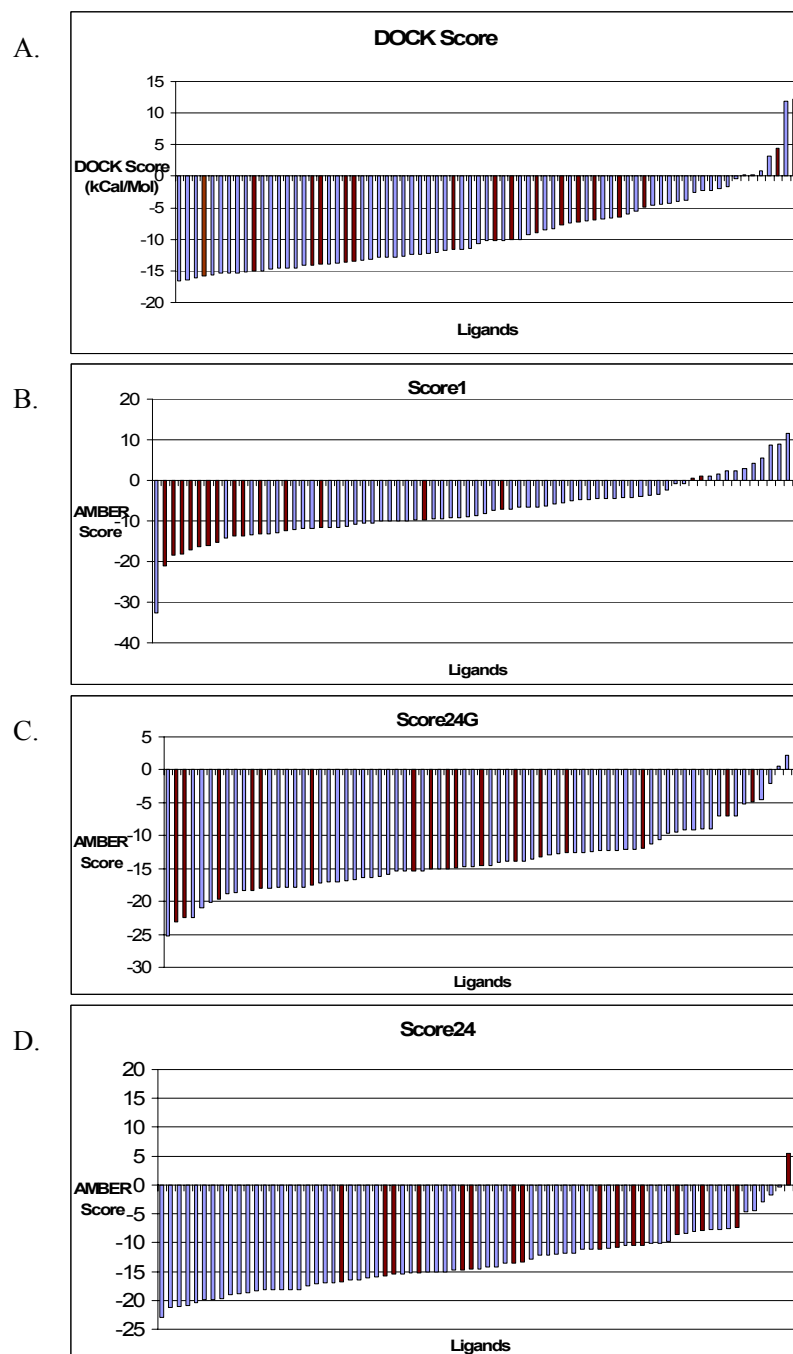
**Table S9 (continued).** Top 100 hits predicted by AMBERDOCK for CCP.

#	STRUCTURE	#	STRUCTURE	#	STRUCTURE	#	STRUCTURE
53		65		77		89	
54		66		78		90	
55		67		79		91	
56		68		80		92	
57		69		81		93	
58		70		82		94	
59		71		83		95	
60		72		84		96	
61		73		85		97	
62		74		86		98	
63		75		87		99	
64		76		88		100	



Description of the keywords: gb=2, Onufriev, Bashford, Case (OBC) variant of GB, with  $\alpha=0.8$ ,  $\beta=0.0$ ,  $\gamma=2.909$ . Similar to igb=2 in AMBER.; gb=5, with  $\alpha=1.0$ ,  $\beta=0.8$ ,  $\gamma=4.85$ ; diel=C, uses constant dielectric; dt, time step, ps; tautp, temperature coupling parameter, in ps; gamma\_ln, collision frequency for Langevin dynamics, in  $\text{ps}^{-1}$ ; temp0, target temperature; gbsa, add a surface-area dependent energy.

**Figure S1.** Flowchart and parameters for AMBERDOCK rescoring.



**Figure S2.** Preliminary scoring protocols for AMBERDOCK. The known L99A/M102Q ligands (blue bars) and decoys (red bars) are plotted versus the respective score (x-axis). A. DOCK score. B. AMBERDOCK *score1* with minimization only. This uses PEOE charges,  $gb=1$ ,  $gbsa=1$ , and moveable residues=(78, 84, 88, 91, 102, 111, 118, 121, 153). C. AMBERDOCK *score24G* with MD and minimization. The scoring protocol includes 3000 steps of MD with a moveable region of 5 Å from the ligand, and  $gb=5$ . D. AMBERDOCK *score24* with MD and minimization. This is the same as *score24G*, except ligand charges were calculated at the AM1-BCC level of theory.

*It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.*

*I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.*

*Alan P. Graves, III*  
\_\_\_\_\_  
Author Signature

*9/27/2007*  
\_\_\_\_\_  
Date