Improving Protein-Small Molecule Structure Predictions with Ensemble Methods, or

Using Computers to Guess How Tiny Things Fit Together

By

Darwin Yu Fu

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry

August 31, 2018

Nashville, Tennessee

Approved:

Professor Jens Meiler

Professor Terry Lybrand

Professor Andes Hess

Professor Tony Capra

DEDICATIONS

I would like to dedicate this work to my family for providing me with this opportunity to delve into a fascinating world of science. To my parents, Cary and Yao, without whose sacrifices and risk-taking I would not be here in this country. To my brother, Daniel, for challenging me and giving me the chance to pass on the lessons I have learned. The three of you have been incredible supportive and keep me reminded of where I came from and who I am.

ACKNOWLEDGMENTS

Research is a team effort both inside and outside of the lab. Though my name is on the front page of this dissertation, the research presented here would not be possible without the help of a significant number of people. To start, I would like to thank my advisor, Dr. Jens Meiler, for his mentorship over the course of my graduate career. I remember the first day in lab when he offered me the choice of working on "small molecule docking" or "protein secondary structure modeling", and I just picked one and ran with it for six years. I admire his ability to make deep insightful comments on a project while maintaining the overall wide picture of research goals. He pushed me to test the limits of existing methodologies, and this research truly would not be possible without his support and guidance. I would also like to thank the rest of my thesis committee: Dr. Terry Lybrand, Dr. Andes Hess, and Dr. Tony Capra for their excellent advice and constructive critique on my work. My qualifying committee members Dr. Jarrod Smith and Dr. Corey Hopkins were tremendous in making sure I had a critical understanding of previous literature. I also like to acknowledge my high school chemistry teacher, Mr. Camacho, for being one of the first to pique my interest in Chemistry, and my undergraduate research mentor, Dr. Dong Xu, for giving me the opportunity to play with protein models on the computer.

Much of this work is funded by a PhRMA Foundation Pre-Doctoral Informatics Fellowship and the Molecular Biophysics Training Grant. I like to thank Dr. Walter Chazin for his organization of the Molecular Biophysics Training Program that introduced me to the breadth of structural biology, and for his helpful pointers on the basketball court. The include collaborative projects would not be have been possible without medicinal chemistry and pharmacology parternships. In particular, I like to thank Dr. Craig Lindsley, Dr. Heidi Hamm, Dr. Thorsten Berg, Dr. Cody Wenthur, Dr. Karl Voigtritter, Dr. Kayla Temple, Dr. Matthew Duvernay, and Dr. Shaun Stauffer for their contributions to applications projects contained in this thesis.

One community I want to acknowledge expressly is RosettaCommons. My first scien-

tific conferences in graduate school was RosettaCon in 2012 and since then I have learned a great deal about programming and modeling from the community. In particular, Andrew Leaver-Fay, Steven Lewis, and Jared Adolf-Bryfogle have taught me quite a few insider tips to Rosetta.

My Meiler lab colleagues, past and present, have been in the trenches with me through the years. Gordon Lemmon, Sam DeLuca, Rocco Moretti, and Steven Combs helped me work through many of my RosettaLigand questions. Jeff Mendenhall, Axel Fischer, David Nannemann, and Jonathan Sheehan helped me with programming tips and scientific discussions. Amanda Duran, Alberto Cisneros, and Alyssa Lokits have been great colleagues in lab and good friends outside lab. I thank you for making the lab such a welcoming place, and helping me celebrate the ups and move past the downs.

The rigors of graduate school would have been tougher if it were not for my friends outside the lab. My soccer/basketball teammates, and tennis partners may not realize it but they played a big role in keeping me sane. I like to acknowledge Mackenzie, Rose, Monika, Carl, Leslie, and Sherri for being great friends, trivia teammates, and graduate school sounding boards. I also wish to express my gratitude to my non-graduate school friends, Abigail, Sherry, and Ari, for confirming that there is indeed life outside of the lab. This seems as good place as any to acknowledge Lucky Bamboo and Corner Asian for keeping me fed with homestyle Chinese food while I'm in Nashville.

Lastly, a few more thank yous for people whose love and support goes way beyond the thesis pages. To Megan, my adventuring partner in life, for all the encouragement until every last word of this dissertation was done. To my brother, for all the things you have taught me about being patient and kind. To my dad, for all the computer knowledge you have given me since the days of when I was just playing with floppy disks in your office. To my mom, for all the times you drove me, both literally and figuratively, to learn about the world. This would not have been possible without you all and I am excited for the future both in life and in science.

TABLE OF CONTENTS

LIST OF FIGURES

SUMMARY

Understanding protein-small molecule interactions is a critical part of exploring protein structure and function. Protein-ligand docking is the computational method aimed at predicting if and how a protein and a small molecule interacts. The work presented in this thesis examines ways to improve upon existing docking methodology through the use of experimental structure-activity relationships and structural ensembles. The algorithms are then applied in collaborative efforts towards structure-based drug discovery.

Chapter 1 is an introductory chapter discussing the background of incorporating experimental information and structural ensembles into protein ligand docking. It contains material from Bender et. al. "Protocols for molecular modeling with Rosetta3 and RosettaScripts", a protocols manuscript for which I was an equally contributing first author. I was responsible for the protein-ligand docking section as well as developing a workshop tutorial on RosettaLigand docking. There is also material from Fu & Meiler "The Predictive Power of Different Types of Experimental Restraints in Small Molecule Docking: A Review", a review for which I was the sole first author. The datasets analyzed are discussed in Appendix A.

Chapter 2 details the development of RosettaLigandEnsemble. It contains materials from Fu & Meiler "RosettaLigandEnsemble: A Small Molecule Ensemble Driven Docking Approach" for which I am the sole first author. I programmed and benchmarked RosettaLigandEnsemble and wrote the manuscript. Appendix B contains the protocol capture accompanying this chapter. The congeneric ligand dataset is discussed in Appendix A.

Chapter 2 is based on an in-prep manuscript regarding the improvement to an automated protein ligand docking server. Further feature additions to the server, located at http://rosie.rosettacommons.org/ligand_docking, is ongoing. I contributed to the server, drafted the manuscript, and will be an equally contributing first author on the manuscript.

Chapter 4 covers the applications of RosettaLigand and RosettaLigandEnsemble in collaborative drug discovery efforts. I made protein modeling and ligand docking models to

support experimental data. The STAT collaboration resulted in a publication "Development of Erasin: A Chromone-Based STAT3 Inhibitor Which Induces Apoptosis in Erlotinib-Resistant Lung Cancer Cells" for which I was a co-author. An additional manuscript from the STAT project is presently under review. The mGlu3 collaboration is no longer being pursued in its existing form. The PAR collaboration is an ongoing collaboration likely to result in a publication at a future time.

Appendix A contains the datasets analyzed in Chapter 1 and 2. There are additional datasets generated as a common lab resources for future benchmark studies.

Appendix C is an ongoing project extending RosettaLigandEnsemble to include ensemble docking with protein mutations. The primary algorithm development and proof of concept study will be submitted as an Arxiv preprint with a peer reviewed manuscript to follow once an extensive benchmark is complete. Additional features added to Rosetta will enable future developments in multitarget virtual screening and ensemble docking for highly dynamic proteins.

Chapter 1

Introduction

1.1   Summary

This chapter introduces protein small molecule docking and provides an overview of its applications in drug discovery. The feedback loop between computational predictions and experimental data is critical step in the modeling pipeline. This chapter discusses methods for feeding wet-lab data into docking simulations, and ways for computational results to inform further wet-lab experimentation.

This chapter contains material published as Bender et. al. "Protocols for molecular modeling with Rosetta3 and RosettaScripts"[1] for which I am an equally contributing co-first author, and Fu & Meiler "Predictive Power of Different Types of Experimental Restraints in Small Molecule Docking: A Review"[2] for which I am the sole first author.

1.2   Proteins and their small molecule partners

It is a truth universally acknowledged, that a single protein in possession of a good binding surface, must be in want of a ligand. A critical aspect of many proteins' functions in nature is the binding of ligands. Ligands range in size from individual ions such as the magnesium in chlorophyll to large biopolymers such as DNA or RNA. Proteins may modify a bound ligand, such as in the case of enzymes, or may need it to function, such as in the case of cofactors or coenzymes. One prominent example of the latter is the role ascorbic acid plays in making collagen, a structural protein that makes up tissue such as tendons, ligaments, and skin. A lack of ascorbic acid, or vitamin C, impairs a hydroxylation step in collagen synthesis resulting in scurvy, a disease marked by joint weakness and skin bleeding [3]. Conversely, some ligands specifically inhibit protein function. Penicillin

binds a protein involved in bacterial cell wall peptide synthesis, resulting in a damaged cell wall and bacterial cell destruction. A whole class of penicillin derivatives have been created with anti-microbial properties via protein inhibition [4]. Noted chemist George Scatchard mused there is such wide variety in the ways proteins interact with ligands that it would be impossible to try to categorize all the forms of protein-ligand binding[5].

As one might imagine, controlled promotion and inhibition of particular proteins have tremendous use in medicine and in correcting aberrant protein function. One class of these protein binding ligands are small molecules, organic compounds larger than ions but smaller than biologics. According to DrugBank, roughly 70% of approved drugs are small molecules that bind reversibly to proteins as biological regulators. Lipinkski's rule of five provides additional general guidelines to these "drug-like" molecules including a molecular weight below 500 Daltons and a limited number of hydrogen bond acceptors and donors. Generally, it is much easier to experimentally determine "if" a small molecule binds to a given protein, then to determine the "how, where, and why". The latter questions require an understanding of molecular structure and the biological process the compound aids or disrupts. The challenge of elucidating protein-ligand structural questions is one of the central aims of computational docking.

1.3   Computational protein - small molecule ligand docking

Protein-ligand docking aims to predict computationally the binding interactions between a protein and a small molecule ligand. This requires a combination of recapitulating the binding pose and quantifying the interaction strength. Docking is an important step in the pipeline for structure based design and discovery of small molecule drugs. Successful prediction of binding position is necessary to delineate critical interactions for improving selectivity and/or efficacy. Docking interrogates how and why a ligand binds to a protein. Tangentially, virtual screening and ligand ranking asks if and how strongly a ligand will bind. Together these techniques have aided advances in lead compound hit discovery

and optimization[6]. One of the earliest examples of ligand "docking" involved building a scaled model of hemoglobin by hand[7]. A first actual computational docking used shape complementarity to dock protein surfaces by describing them as knobs and holes. The study successfully matched two dimers of hemologin but could not dock a trypsin and an inhibitor because the potential combinations of the two required too much computing capacity[8].

Computational ligand docking exist in conjunction with traditional wet-lab experiments. As Richard Hamming once noted, "The purpose of computing is insight, not numbers". Computer models provide hypotheses that must be validated experimentally. The results of these tests are then fed back to create an iterative cycle of model improvement. The attraction of computer modeling is the direction it provides towards the efficient use of wet-lab resources.

1.4 Ligand docking software

Popular docking algorithms such as AutoDOCK[9], DOCK[10], GLIDE[11], GOLD[12], and RosettaLigand[13], have diverse methods for representing, sampling, and scoring the molecular interface. Each sampling and scoring setup has its own advantages and disadvantages with regards to computational cost and accuracy. Many of these pros and cons are system or use-case dependent. RosettaLigand is the core algorithm used in this thesis and will be detailed separately.

For sampling, AutoDOCK[9] and GOLD[12] uses genetic algorithms that generate a series of trial conformations. Each set of trial conformations are allowed to change before the ones with the best binding energy is selected in a manner akin to natural selection. Incremental growth algorithms such as DOCK[10] and GLIDE[11] break the ligand down into fragments and start placement at an "anchor". The molecule is then reconstructed piece by piece to fill the binding pocket. Monte Carlo methods such as RosettaLigand[13] make a random perturbation to the protein-ligand model in each step and then accepts the

move if it has a more favorable score, or resets it if the move generates a less favorable score. To avoid being trapped in a local minimum, there is a slight probability that a less favorable move will be accepted based on the Metropolis criterion.

Scoring functions are divided into three categories: force-field, empirical, and knowledge-based[14]. Force-field scoring is often referred to as "physics based" but in reality, docking algorithms generally simplify calculations or use empirical parameters to reduce the computational cost. Force-field based methods such as MM-PBSA and MM-GBSA may be used to rescore select docking models previously produced with other score terms[15]. AutoDOCK[9], DOCK[10] both began using an AMBER force field before adding empirical scoring terms. Empirical score functions generally sum up individual scoring terms over pairs of specific interactions. For example, the ChemScore function used in GOLD[12] and GLIDE[11] evaluates all hydrogen bonds and assign favorable scores to ones within a certain distance and angle range. RosettaLigand[13] uses a knowledge-based scoring function for many aspects of its protein structure scoring. The scoring function is based on statistics from the Protein Data Bank regarding the most common distances and angles for protein features, and awards the best scores to models that match previous data.

1.5 RosettaLigand

RosettaLigand is a protein-small molecule docking application within the Rosetta Macromolecular Modeling framework. It is a Metropolis Monte Carlo sampling method paired with a knowledge based scoring function designed to consider both protein and ligand flexibility[16, 17]. RosettaLigand uses a two phase docking approach: a low resolution phase of rapid sampling based on shape complementarity followed by a high resolution phase of sidechain repacking and backbone minimizing. Ligand flexibility is modeled through pre-generated conformations[18]. Using default settings, RosettaLigand accounts for significantly more protein flexibility than comparable methods. As such, the two phase method allows for many ligand placements to be sampled without being bogged down in

the computationally intensive task of optimizing a fully flexible protein structure.

The low resolution phase scores binding poses based on a pre-generated Van der Waals attraction-repulsion grid. A band of unfavorable repulsive interactions and a band of favorable attractive interactions are placed at set distances around a rigid receptor model. Ligand atoms outside of this scoring shell are considered neutral scoring. Scoring grids can also be pre-computed based on hydrogen bonding and other non-Van der Waals interactions[18]. The high resolution phase scores models using the Rosetta energy function. As previously discussed, the energy function uses PDB-derived knowledge-based potentials for scoring protein features such as Ramachandran angles and sidechain rotamers. Pairwise interactions such as Van der Waals attraction-repulsion, electrostatic interactions, hydrogen bonding, and disulfide bridges are weighted to units of kilocalories per mole. Many of Rosetta's interaction potentials are derived from the CHARMM forcefield. Further details of the Rosetta energy function can be found in Alford et. al.[19]

Features for RosettaLigand are selected and customized through an XML interface[20], making it easier to add new docking capabilities. A sample of previous RosettaLigand development include docking with explicit interface water molecules[21], protein-ligand interface design [22], and rapid screening of ligand libraries[18]. These protocols have driven an improvement in Rosetta's docking accuracy in simpler cases as well as expand RosettaLigand's applicability in more complex systems. For example, explicit water docking demonstrated 56 percent recovery for failed docking cases across a diverse benchmark of 341 structures [21]. RosettaLigand's performance in community docking benchmarks is discussed in the following section.

1.6   Community assessments of docking

Ligand docking is commonly assessed by determining if docking programs can predict the binding mode given an interacting pair of protein and ligand. Depending on the availability of data, structures of existing homologs or a priori knowledge of the binding

site may also be provided. Two related challenges are whether or not programs can rank a series of binding ligands with a mutual target, or determine if a particular ligand binds to the given target.

Community assessments of docking software have generally displayed success in recovering near-native binding poses. Davis et. al. found that accurate binding poses were found for all targets in a GlaxoSmithKline compound collection, but the overall success rate varied dramatically among systems. Furthermore, no algorithm consistently outperformed the others across all systems. RosettaLigand generated native-like poses for at least forty percent of ligands in each target set. RosettaLigand had difficulty in cases with tight binding deep pockets as the pre-computed ligand conformers could not account for the precise binding geometry needed[23].

The CSAR 2012 benchmark demonstrated features such as protein structure minimization, histidine tautomeric states correction, pre-generated small molecule conformations, native small molecule training, and substructure based restraints correlated positively with docking success. However, binding affinity prediction and relative ranking of active small molecules remains the most challenging aspect in the field and during this experiment in particular[24]. In the 2015 D3R Grand Challenge involving blinded docking of HSP90 and MAP4K4 ligands, roughly half of submitted predictions had a near-native model as the best prediction. Not surprisingly, the most successful workflows superpositioned targets onto similar ligands instead of sampling large, unrestricted binding space. One observeration of note was that the same docking software used by different groups produced varied results depending on the exact parameters for ligand placement and scoring. RosettaLigand was such a case where performance differed significantly based on the initial alignment method and the allowed sampling volume [25]. The second iteration of the D3R Grand Challenge affirmed the broad results of the 2015 benchmark and highlighted the need for regular blinded assessments to evaluate development[26].

1.7   The use of experimental data in docking

A common theme of the previously discussed docking assessments was the benefit afforded by relevant experimental data. Experimental data may be straight forward in application such as in the case of receptor structures determined by X-ray crystallography; these structures may be used directly as docking targets. Some experimental data may be more "fuzzy" as in the case of protein-ligand contact points determined via NMR spectroscopy. With the proper integrative docking algorithms, these interacting points serve as distance restraints that can be incorporated as part of the sampling and/or scoring process. Restraints limit the ways a protein-ligand interface can be constructed, hence reducing the sampling the sampling complexity and improving the scoring discrimination[24, 16, 25, 27]. The potential for such soft restraints in hybrid/integrative methods have already been reported for other aspects of protein modeling, such as the use of cryo-electron microscopy[28] or electron paramagnetic resonance[29] restraints in protein structure prediction.

Experimental information for guiding ligand docking can be classified as protein receptor structure-based, small molecule ligand-based, or protein-ligand interface-based. Protein receptor structure-based data provides information about the protein target in the form of an observed conformation, or knowledge regarding the ligand binding site. Conversely, small molecule ligand-based indicate the ligand components responsible for interacting with the target. Protein-ligand interface-based measurements are a combination of the two and directly identifies a specific protein-small molecule interaction. Although the same experimental technique (ex: NMR spectroscopy) may be used to generate any of the three kinds of restraints, the computational guidance provided by each data type differs significantly. The following sections discusses the overall value of experimental restraints, the strengths and drawbacks of each form of experimental data, and examples of programs/methods that apply each data type.

1.8   The power of different restraint types

In order to demonstrate the different power of the various restraint types, a simulated re-strained docking benchmark was conducted using RosettaLigand[20] on the PDBBind Core Subset described in Appendix A.2. Protein receptor structure-based data were represented by restraining three randomly chosen binding pocket residues contact any small molecule heavy atom. This is analogous to experimental data that identifies residues whose shifts change in an NMR experiment upon ligand addition, or whose mutations abolish ligand binding. Small molecule ligand-based restraints are simulated by requiring three particular small molecule heavy atoms contact the protein. This is akin to structure-activity rela-tionships that show binding affinity changes when certain functional groups are swapped out. Protein-ligand interface-based restraints are created by enforcing contacts on three randomly chosen pairs of interacting atoms between the receptor and the small molecule. These types of restraints are derived from experiments such as a double mutant cycle that correlate a specific protein-ligand contact point. A contact was defined as an interatomic distance less than 4 , which includes most commonly observed molecular interactions[30]. For each test, the small molecule was initially subject to a random reorientation and transla-tion within a 5 sphere. An additional test was conducted using minimal initial perturbation as a representation of using the binding mode of a homologous protein-ligand complex as a guide to initial placement. This case arises in applications where the a ligand highly similar to the ligand of interest has a known binding mode with the given protein receptor. 2500 docking trajectories were completed for each protein-small molecule test case under each restraint condition. The models were analyzed for percentage of native-like small molecule binding modes using a 2.0 RMSD cutoff.

Figure 1.1 shows the distribution of sampling success rates across test systems for each of the restraint conditions. The largest improvements are seen in docking with interface-based restraints and in using molecular similarity to restrict the starting position. This makes intuitive sense as both restricts the small molecule rotational orientation in additional
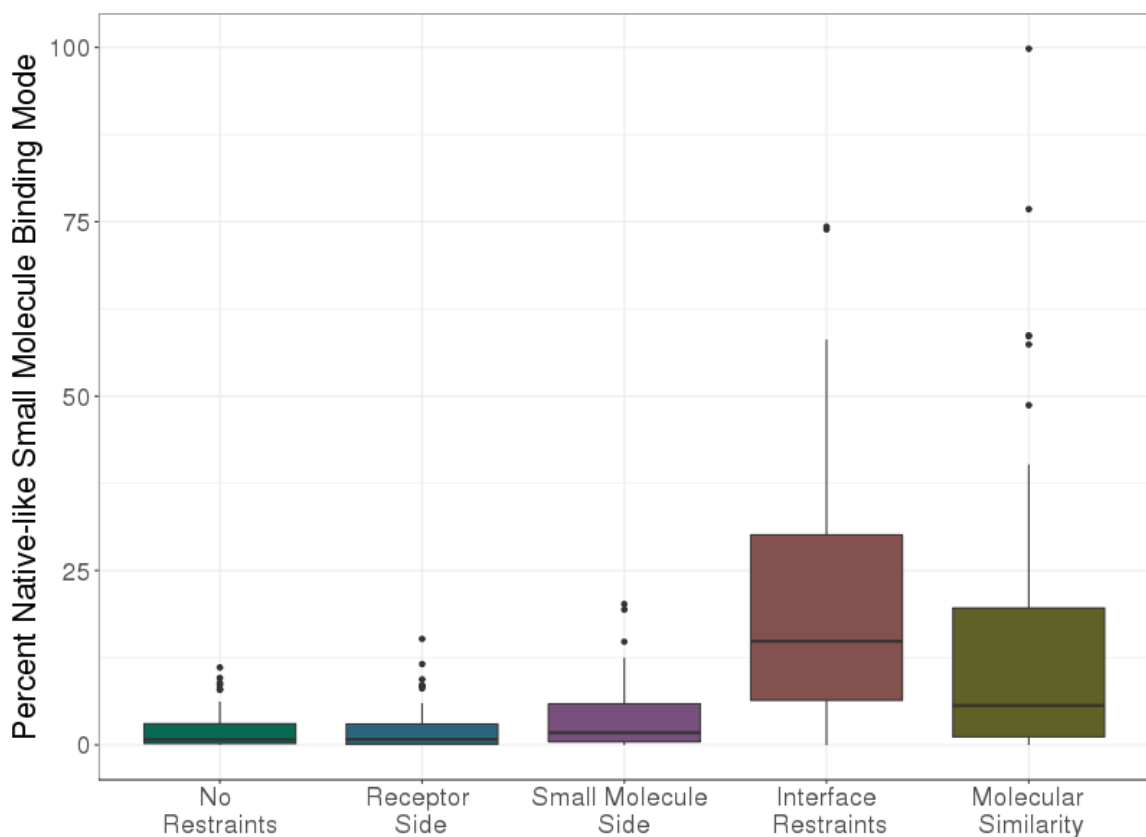
Figure 1.1: Boxplots of simulated restrained docking sampling efficiency. Each boxplots show distribution of percent native-like binding modes observed across the PDBBind Core subset for each restraint condition.

to its translational location in the binding pocket. Certain interface-based techniques, such as INPHARMA, specifically work by determining the relative orientation of two similar small molecules. Reinforcing interatomic distances with a protein point and a ligand point restricts both the protein side-chain and small molecule conformational flexibility. Figure 1.2 provides examples of each type of restraint. Molecular similarity is discussed later in this chapter as it is a general assumption about protein-ligand families.

Technique	Information
X-Ray Crystallography	Static model of apo protein or protein-small molecule complex
Nuclear Magnetic Resonance	Ensemble of protein models
Chemical Shift Perturbations	Protein residue signal changes with varying small molecule concentrations
Nuclear Overhauser Effect (Intra-protein)	Protein structural changes upon small molecule binding
HD-Exchange Mass Spectrometry	Protein amide hydrogens buried by addition of small molecule
Isothermal Titration Calorimetry	Binding affinity changes due to protein mutations or small molecule alterations
Structure − Activity Relationships (DoMCoSAR)	Correlation between small molecule modifications and binding affinity
Consensus Molecular Shapes (SABRE)	3D structural overlay of multiple active small molecules
Protein-Mutation/Ligand-Modification Coupling	Specific contacts based on correlation of protein and ligand modifications ("Double Mutant Cycle")
Inter-ligand NOEs (INPHARMA)	Relative orientations of two small molecule binders
Protein-Ligand NOEs	Distances between protein-small molecule contact points
Ligand Molecular Similarity (HybridDock\, LigBEnD)	Binding mode of a related small molecule to the same protein target

Figure 1.2: Receptor-based (blue), small molecule-based (red), and interface-based (green) experimental data. Example programs for particular methods are given in parentheses.

1.9   Protein receptor structure-based data

Structure-based, also referred to as receptor-based data, are derived from observed changes or effects on the protein alone. Protein structures in the absence of (apo) or in complex with the small molecule (holo) determined via X-ray crystallography are the most straightforward form of structure-based data. Small molecules can be directly docked into receptor crystal structures or, if such structures are unavailable, into homology models. Docking into holo crystal structures is generally more accurate than docking into apo crystal structures or comparative models[31]. In testing a nitroreductase protein-ligand target from CASP 11, Huwe et. al. found few dockings to comparative model structures that were superior to docking to the experimental crystal structure. However, the comparative

model docking managed to capture specific contacts 72.7% of the time[32]. Bordogna et. al. found a Spearmans correlation coefficient of 0.66 between RMSD accuracy of the comparative model and the accuracy of the docking simulation for a diverse test set[33]. In high-throughput docking, or virtual screening, applications, comparative models are capable of similar enrichment rates as their crystal structure template counterparts[34, 35]. A common theme across the assessments is that traditional measures of comparative modeling ease, such as sequence similarity between template and target, does not correlate with subsequent docking success. The success rate for docking into homology models can be improved by up to 70 percent by using holo experimental templates crystallized with small molecules of similar chemotypes[36]. Careful validation of the input crystal structure, particularly in regards to proper orientation and placement of the small molecule, should be performed prior to using the structure in computational drug discovery efforts. Any modifications to the target protein in the crystallization process, including biologically irrelevant mutations or inserted constructs, should also be considered[37].

Receptor structures may also be derived from nuclear magnetic resonance (NMR) spectroscopy. An ensemble of conformations is generally provided to capture the flexibility observed in structures obtained by NMR spectroscopy. Alternatively, NMR spectroscopy may be utilized to obtain information on protein-small molecule interface contacts. Chemical shift perturbations[38] are observed for specific residues upon small molecule binding, while intra protein Nuclear Overhauser Effect (NOE)[39, 40, 41] reflect structural changes within the protein. Distance restraints derived from these two sources are based on the assumption that changes are due to interactions with the small molecule. Protein-focused methods can help define the receptor binding pocket but do not necessarily give information on the small molecule binding mode. This type of information generally translate to a restraint favoring small molecule positons that are within a certain distance of the contact residue[42]. Orts et. al. demonstrated the use of protein-mediated NOE data for two competitively binding small molecules as a post-docking scoring filter that can improve ac-

curacy by two orders of magnitude[40]. Onila et. al. extended this method to directly use NMR data during docking by simultaneously optimizing poses for both small molecules, which improved docking in a test set of weakly bound cAMP-dependent protein kinase complexes. However, the results were highly dependent on obtaining proper orientation of the protein side chains[39]. Cala et. al. reviews further experimental details for NMR characterization of protein-small molecule contacts[42].

Other methods for studying localizing small molecule binding site interactions include hydrogen/deuterium exchange mass spectrometry (HDX-MS) and isothermal titration calorimetry (ITC) used in conjunction with mutagenesis. HDX-MS relies on the different exchange rates for exposed versus buried amide hydrogen atoms to identify protein residues covered up by small molecule binders[43]. Mouchlis et. al. uses HDX-MS on protein backbone amides in conjunction with docking to determine binding modes of phospholipase A2 inhibitors[44]. ITC is used to measure the thermodynamic components of binding affinity: enthalpy and entropy. Structural information is inferred from binding affinity changes following protein mutagenesis or small molecule modification. It is generally assumed that such changes is due to the impact of the alteration on small molecule binding[45].

1.10   Small molecule ligand-based data

Small molecule-based information takes advantage of binding data through quantitative structure-activity relationships (QSAR). QSAR is traditionally part of virtual screening applications when no receptor information is available. These models generally make use of 2D small molecule property descriptors or 3D small molecule shape fitting without constructing a receptor model. Certain docking algorithms can incorporate comparisons of known small molecule binders in generating putative binding modes. SABRE is a method that generates a consensus molecular shape density function from multiple bioactive small molecules. Candidate molecules are then shape fitted using chemical substructures as op-

posed to the entire molecule at once[46]. Adherence to experimental SAR can also be used as a filtering step, though this is tricky as scoring functions generally do not rank order compounds well. DoMCoSAR is a docking algorithm that selects the most commonly observed binding modes and utilizes correlation with SAR to guide final model selection[47]. Small molecule conformational shape fitting can also be achieved with transfer NOE as demonstrated in the design of a flexible macrocyclic inhibitor[48].

Ligand based pharmacophore models are an extension of molecular shape fitting by identifying chemical commonalities among binders of a given receptor. Known ligand binders of a receptor are aligned as flexible conformers and common features are identified in 3D space to identify the pharmacophore. This approach can also be performed with protein side chains to create receptor side pharmacophores[49].One such method PharmDock converts the receptor and ligand to hydrogen bonding and hydrophobic pharmacophores before using an alignment algorithm to match pairs. On a test of the PDBBind Core Set, PharmDock identified a native-like top model 56% of the time when native conformations were used but only 37% when using Omega generated conformers, signifying the importance of input pharmacophore conformations[50]. Pharmacophores can be generated in combination with molecular dynamics to generate an ensemble of binding pocket models as shown in a development study of ligands for a highly flexible sulfotransferase binding pocket[51]. Yang et. al. further discuss both ligand and structure-based pharmacophore modeling along with potential challenges such as dataset construction, molecular alignment, and feature selection[49].

Use of SAR and multiple active small molecules in docking simulations is an exciting area of research. In particular, activity cliffs, highly similar compounds with orders-of-magnitude differences in potency, provide powerful SAR information. New methodologies that utilize the experimental knowledge provided by SARs around activity cliffs can guide the creation of additional structural analogues[52]. The current release of the ChEMBL database contains over 13 million activities recorded against over 10,000 targets[53]. Fur-

thermore, there are system specific small molecule SAR databases for common drug discovery targets such as GLASS for GPCRs[54] and KLIFS for kinases[55]. One potential way to incorporate this data is to use an ensemble docking method that can simultaneously optimize multiple protein-small molecule complexes and correlate their calculated scores. MLSD is an extension of AutoDock4 that allows the simultaneous optimization of multiple small molecule fragments, though it is restricted to molecules that concurrently bind to the same target[56]. Mass spectrometry of protein unfolding can also be used to examine multiple ligand bindings and their combined interactions on protein stability in the gas phase, though further methods are necessary to translate this to binding modes[57].

In a reverse modality, small molecule information can also be used to generate distance-dependent pair potentials for protein comparative modeling. The MOBILE program docks small molecules into a starting ensemble average of homology models, and then generates restraints for subsequent rounds of model refinement. When used in combination with MODELLER, MOBILE produced native-like binding pocket geometries in 70% of test cases, improving results in 60% of cases compared to restraint free comparative modeling[58].

1.11   Protein-ligand interface-based data

Protein-small molecule interactions are the most powerful combination of the protein and small molecule information as it can identify specific contact points that can be used in determining both location and orientation. In testing small molecule docking into G-protein coupled receptors, Nguyen et. al. demonstrated that sampling efficiency can increase by an order of magnitude for every ten known protein-small molecule contacts. The gain was even greater when utilizing more detailed information such as a specific ionic interaction translated as a 3.0 distance restraint[59].

A protein-small molecule double mutant cycle analysis identifies interactions by comparing ITC binding data of a single protein mutation, a single small molecule functional

group substitution, and both simultaneously. A substantial non-additive interaction energy change is evidence for a direct interaction. A collection of these pairwise interactions can be used to derive protein-small molecule distance constraints to incorporate into the energy gradient docking grid. Roisman et. al. probed 100 potential pairwise interactions in an interferon-receptor complex and identified five significant interactions. Docking simulations were run until a converged model was generated satisfying the five restraints[60]. Blum et. al. utilized double mutant cycle analysis with a nicotine analogue and an acetylcholine receptor backbone amide substitution to show a hydrogen bond interaction[61]. Similar success have been obtained with double mutant cycle alanine scanning for a yeast Ste2p GPCR[62], and with an allosteric binding site on a hM1 muscarinic receptor complex[63]. Compared to the traditional single site-directed mutagenesis method, this approach has the potential to differentiate the impact on small molecule binding from disrupting favorable interactions vs. protein stability.

Another method of directly determining intermolecular distance restraints is with protein-ligand NOEs. However, this is generally limited by the need to assign resonances for the protein-small molecule complex. One alternative that relies on matching only small molecule resonance assignments was developed by Constantine et. al [64]. NMR NOE experiments can also be used to derive relative orientations between two weakly binding, competitive small molecules. The INPHARMA technique relies on the transfer of NOEs between the two small molecules and a common receptor target. This in combination with a crystal structure of one protein-small molecule complex can be used in determining the binding of a related small molecule series without assuming binding in a similar fashion[65].

Small molecule similarity is a type of interface restraint based on the known binding mode between a related small molecule and the given protein target. The binding mode of the related molecule can be used as a guide in placement and orientation during docking simulations. HybridDock is one approach that augments docking with molecular similarity

by generating possible binding modes using existing co-crystallized molecules. This significantly improved both the binding energy correlation and native binding mode recovery in a CSAR 2013-2014 test[66]. LigBEnD is a similar approach that uses the co-crystallized molecule to generate an atomic property force field. Scoring models with this ligand-based force field correctly predicted 30 out of 36 compounds in the D3R docking challenge[67]. Related binding pockets may be found even among proteins of distinct global folds and evolutionary history. An analysis of potential enzyme drug targets and evolutionarily distant proteins in the PDB found similar binding pockets with different global folds in 61%, 10%, and 61% of kinases, phosphatases, and proteases respectively[68].

1.12 Similar binding of similar ligands

One potent form of experimental restraint as shown in figure 1.1 was molecular similarity, the assumption that similar ligands will bind in similar fashion to a given target. Although the relative orientation can be experimentally tested, it is often assumed as a starting point. The common molecular scaffold is presumed to make similar interactions with the binding pocket, while peripheral functional group modifications create different contacts that explain SARs. Previous analysis of 206 protein-small molecule structure pairs observed similar small molecule binding modes in ninety percent of related structures. Binding similarity was defined as having an optimized small molecule shape Tanimoto of greater than 0.8. The receptors in those cases exhibited very similar backbone structures with the primary differences due to side chain conformation or water architecture[69]. An examination of scaffold building pairs found that in 41 out of 297 cases the binding mode changed upon chemical elaboration of a scaffold[70].

Using a subset of the PDBBind Refined Set, described in Appendix A.3, an analysis is done to show the feasibility of the similarity approach across a broad dataset. The comparison of 7298 ligand pairs from 366 targets, shown in Figure fig:Review-Fig3, demonstrated a significant decrease in RMSD of the common scaffold as the Tanimoto similarity increased.

Disparities in binding mode decreases significantly for Tanimoto similarities above 0.6.

Tanimoto Range	Total Pairs	% Under 2Å nRMSD
0.0 - 0.2	2038	31.31
0.2 - 0.4	2106	59.64
0.4 - 0.6	1198	81.55
0.6 - 0.8	1062	91.53
0.8 - 1.0	894	94.18
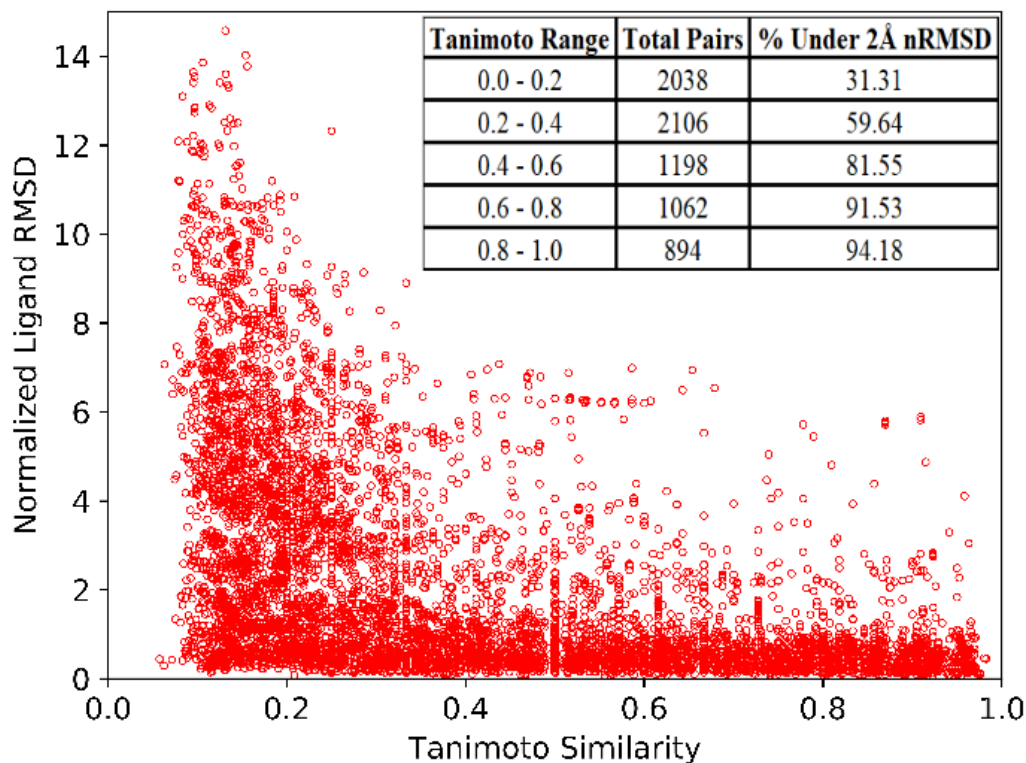
Figure 1.3: Pairwise small molecule scaffold nRMSD vs. Tanimoto similarity Inset: Number of pairs and percentage under nRMSD cutoff for each Tanimoto range.

The dataset included 548 pairs where the common substructure is equivalent in size to the smaller molecule. A small molecule binding mode change (nRMSD >2.0) was observed in 52 cases (9.5%), slightly lower than the percent changed based on volume overlap comparison used by Malhotra and Karanicolas[70]. Although these cases present a challenge to using molecular similarity docking, it is generally possible to recover the correct binding mode in these scenarios. In particular, a number of properties such as pocket volume and molecular weight of the smaller molecule can be used as predictors of when these exceptions occur. The much more challenging systems contain similar ligands presenting in completely opposite binding modes. These exceptions are likely to lead to docking failures when assuming molecular similarity and are discussed in more detail in

the following sections.

1.13 Significantly different binding modes observed in similar ligands

In eleven small molecule pairs, highly similar small molecules (Tanimoto >0.7) exhibited significantly different binding modes (nRMSD >5.0). In one particular example, a series of diflunisal derivatives exhibited two opposite orientations when bound to transthyretin, a protein involved in amyloidogenesis. The lead compound diflunisal was found to bind in a forward and a reverse orientation. More interestingly, a meta-difluoro derivative (Figure 1.4, left) was found exclusively in the reverse binding mode while an ortho-difluoro derivative (Figure 1.4, right) was found exclusively in the forward binding mode[71].
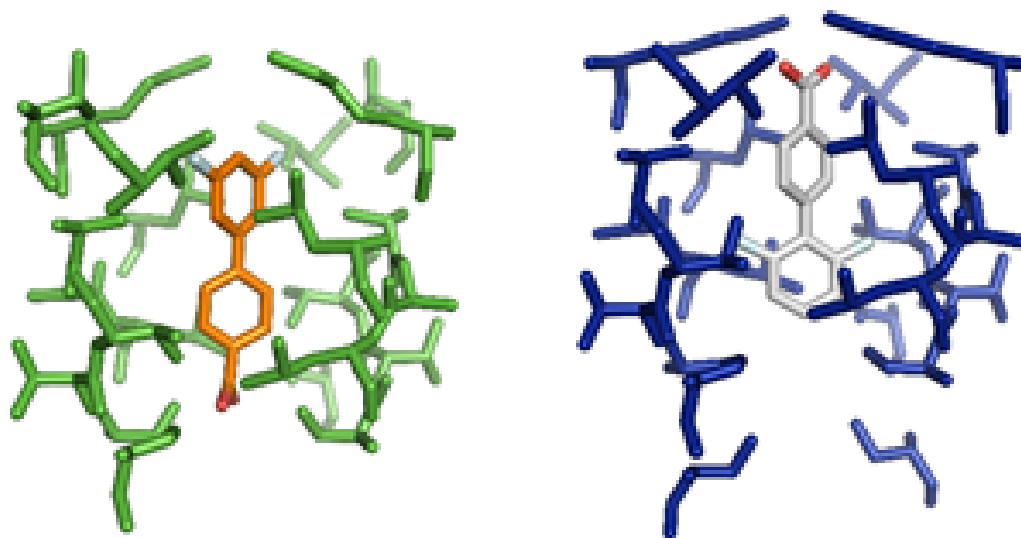
Figure 1.4: meta-difluoro diflunisal derivative (left, PDB: 2B9A) and ortho difluoro diflunisal derivative (right, PDB: 2F7I) bound to transthyretin.

Another counter-example in the dataset is the binding of bile acids, taurocholate and cholate, to Campylobacter jejuni CmeR regulator protein. The two compounds differ by a distal anionic group but are found bound in anti-parallel orientations as shown in Figure 1.5. The two compounds share the same volume of the binding pocket and is suggested to

18

interact similarly with a previously identified glycerol binding site. Furthermore, the pocket is highlighted by a large hydrophobic tunnel with numerous mini-pockets, suggesting a reason as to why it is capable of binding diverse small molecules in diverse fashions[72].
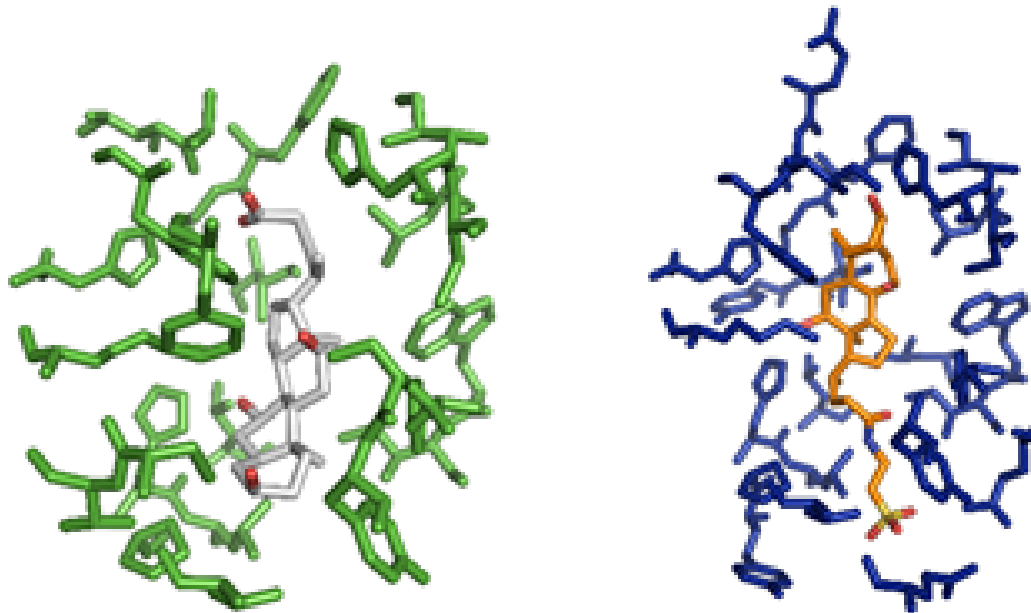
Figure 1.5: cholate (left, PDB: 3QPS) and taurocholate (right, PDB: 3QQA) bound to CmeR regulator protein.

A number of notable exceptions can be found in literature as well. Structure-based design of an influenza neuraminidase inhibitor series showed up to 180 degree variation in the orientation of a central five member ring. Potent analogues were only found for the congeneric series that bound in the same orientation with consistent SAR[73]. A study of dipeptidyl peptidase IV inhibitors showed chemically similar small molecules with different distal aromatic substitutions and placements bind in distinct orientations. The substituted phenyl ring made pi-pi interactions but with distinct residues in the different cases[74]. Another common exception in systems such as HIV-1-Protease involve inhibitors bound in two approximately symmetrical orientations[75]. Kim et. al. discusses a number of other exceptions, such as dihydrofolate reductase and cytochrome c peroxidase small molecules, that were identified through examination of outliers left out

when constructed QSAR datasets. It should be noted that some of the exceptions involve conformational changes in the distal parts of the small molecule while the main chemical scaffold remains aligned[76]. It may also be possible in these QSAR datasets that similar small molecules reside in different conformations of a flexible protein binding pocket. In such cases, the small molecule orientations remain constant but the protein-small molecule contacts change[77].

Based on the PDBBind refined set survey, these exceptions are fairly uncommon. Some features frequently seen in these exceptions include nearly symmetrical molecules, large binding pockets allowing multiple orientations, and distal groups capable of making favorable interactions with different residues in the binding pocket. Unfortunately, there are no currently known small molecule or receptor structure factors to distinguish exceptions from regular binders.

Although a similarity based approach towards docking or screening with atoms aligned by identity may not work in these particular cases, there may be remedies using molecular properties. A docking method utilizing pharmacophores with properties such as partial charge or hydrogen bond donor/acceptors can alleviate this problem. Ph4Dock is an example where atoms are represented as electrical charge centers without consideration for identity[78]. Furthermore, similar contact residues are often observed in these situations allowing for productive suggestions of pairwise interaction validation experiments such as double mutant cycles.

1.14 The use of structural ensembles

Structural ensembles take advantage of binding similarities by the simultaneous consideration of multiple structures for the protein and/or ligand. An ensemble can be used to represent different conformations of the same molecule, or a group of related molecules. Protein receptor ensembles are frequently used to account for receptor flexibility. Rueda et. al. demonstrated improvement of cross-docking results using binding site ensembles to

represent protein flexibility[79]. In particular, ensembles of two or more proteins, enhanced for proteins co-crystallized with chemically similar small molecules, performed better on average than single docking or randomly enumerated ensembles[80]. These conformational ensembles can also be derived computationally using the relaxed complex scheme (RCS), a series of molecular dynamics simulations to pre-generate low energy conformations[81]. Experimental data can then filter the conformations to avoid docking efficiency decrease stemming from having a large number of ensembles structures[82]. Sinko et. al. and Feixas et. al. further discuss RCS and other experimental methods to account for protein flexibility in drug design applications[83, 84].

One area of further development would be ensemble methods to work with protein mutants rather than just protein conformations. Such an algorithm would allow for simultaneous docking or screening against multiple targets of biological relevance. This could be beneficial in targeting multiple mutants with a single compound, or in targeting dual receptors as a replacement for combination therapy. Anighoro et. al. demonstrated the applicability of dual inhibitor screening for Hsp90 and B-Raf inhibitors, though the computation was performed independently rather than in conjunction[85]. The related multiple ligand simultaneous docking strategy, where fragments are docked individually before chemically linked, was used to find an inhibitor of STAT3[86]. An ensemble screening method would score the potential ligands against multiple targets at the same time rather than as a post-screening analysis.

The work in the remaining chapters develop a number of algorithms necessary to use molecular similarity and structural ensembles as an aid to protein-ligand docking. Chapter 2 covers the RosettaLigandEnsemble algorithm that emphasizes overlapping binding modes of similar ligands. Appendix C shows a new docking modality allowing for the simultaneous consideration of protein and ligand structural ensembles. Chapter 4 applies these methods for small molecule discovery applications.

Chapter 2

RosettaLigandEnsemble

2.1 Summary

This chapter discusses the creation and testing of a new docking algorithm within the RosettaLigand framework. RosettaLigandEnsemble simultaneously docks an ensemble of small molecules into a single flexible target receptor. The new method shows a significant improvement in sampling efficiency over the previously existing RosettaLigand method. Ligand ensemble docking is a novel approach not previously available in RosettaLigand or any other popular ligand docking program. The text of this chapter contains material published as Fu & Meiler "RosettaLigandEnsemble: A Small Molecule Ensemble Driven Docking Approach" for which I am the sole first author. I was responsible for developing the algorithm, writing the code, and benchmarking the new docking features.

2.2 Introduction

2.2.1 Ligand docking and structure-based drug discovery

Structure-based drug discovery and optimization is a critical technique at the intersection of pharmacology and structural biology. Structure-based computer-aided drug discovery is a powerful way to create hypothesis on ligand binding poses and specific critical protein/ligand interactions that guide the design of improved small molecules[87]. These hypothesis can be tested by a variety of experimental approaches including fluorescence binding studies, calorimetric measurements, NMR spectroscopic studies, or X-ray crystallography. Experimental validation often compares multiple ligands with the wild-type protein or a mutant target[88]. For computer aided drug discovery to maximize its im-

pact on drug discovery, it is necessary for computational ligand docking methodologies to effectively identify correct protein-ligand binding positions.

Structure-activity relationships (SARs) refer to differences in binding affinity or biological efficacy following chemical scaffold derivatizations. Medicinal chemistry makes use of such minor modifications to optimize lead compounds for desired affinity and other pharmacological properties. This creates a massive wealth of SAR data on related ligands for a single protein target. The PubChem database alone contains over 200 million biological activities measurements on approximately 10,000 protein targets[89]. BindingDB specifically organizes a portion of its database into collections of congeneric ligands with at least one co-crystallized with the common protein target[90]. It is generally expected that highly similar ligands form similar interactions when binding to the same target[2]. We hypothesize that a docking algorithm that leverages this information can eliminate a portion of false positive binding poses, i.e. poses that score well but are incorrect.

2.2.2   Inconsistent performance of existing protein-ligand docking tools

RosettaLigand[17, 16], a small docking tool within the Rosetta structural biology modeling software suite[1], is one of several algorithms developed for this purpose in the last few decades. AutoDock[9], DOCK[10], and Glide[11] are other popular methods, all of which differ on both sampling and scoring technique. Performance of these docking tools are not always consistent across systems. A 2013 docking study using the PDBBind dataset evaluated scoring functions for decoy discrimination and scoring correlation. The success rate for identifying correct binding modes from decoys was significantly higher than for discerning weak, middle, and strong binders within a related ligand series[27]. Similar results were obtained in the 2012 Community Structure Activity Resource (CSAR) evaluation, which found that even when docking software was able to recover correct binding poses for a given ligand, few could consistently rank order active ligands[24]. The recent D3R Grand Challenge reaffirmed these findings and noted that docking performance var-

ied even within the same congeneric series. In addition, the overall success of a docking method was dependent on its preparatory workflow[25]. This performance gap between docking and ranking is likely due to the steep energy landscape observed near native binding modes for high affinity protein-ligand complexes. Small perturbations in these regions generally resulted in drastic scoring changes[91].

2.2.3 Use of structure ensembles in docking

Ensemble methods have traditionally been independently approached from the protein and ligand sides. Protein ensembles are a common way of capturing conformational diversity during rigid receptor docking simulations. This need for a structure ensemble can be due to the inherent flexibility of the protein (conformational selection) and/or due to an induced fit effect upon ligand binding. Protein structural ensembles can be generated from experimental determination such as NMR, or through computational methods such as molecular dynamics. One such preparation is the relaxed complex scheme that generates a set of receptor targets for docking[81]. To emulate induced fit with ligand binding, Glide docking can be used to convert all interface residues to alanine to allow for sampling the binding pocket without bias from initial sidechain orientations[92]. For scoring purposes, protein ensembles can be handled by an average energy grid that scores over the ensemble[93], or by using a selection method to identify a single template mid-simulation[94]. Feixas et. al. and Sinko et. al. further reviews the use of multiple receptor structures in drug discovery and design[84, 83].

Ligand structural ensembles are used to represent both ligand conformations and pharmacophore information from multiple ligands. Molecular mechanics or fragment based sampling can be used to generate conformations prior to docking[95]. Hybrid methods incorporate information from multiple ligands to better position a given target. For example, HybridDock performs pre-docking alignment via pharmacophore matching with similar molecules[66]. However, these methods require related co-crystal structures to be readily

applicable.

It has been observed that using well-chosen structural ensembles is advantageous over docking with a single structure, particularly when ensemble proteins are co-crystallized with molecules of similar chemical structure[96, 97]. In this chapter, we developed a two-stage algorithm for ensemble docking of multiple related ligands into a single protein structure.

2.2.4   Incorporating ligand ensemble docking into RosettaLigand

RosettaLigand models protein-ligand interactions with full ligand and protein binding pocket flexibility. This is achieved with pre-generated ligand conformations and protein side-chain rotamer libraries[17, 16]. RosettaLigand is currently capable of docking multiple ligands simultaneously, but only in the sense that they bind the protein jointly (e.g. a small molecule together with a key bridging water molecule or a co-factor with metal ion bound)[21]. Here, we have extended RosettaLigand to RosettaLigandEnsemble (RLE), an algorithm that can identify a binding mode favorable to a superimposed ensemble of congeneric ligands. This allows users to simultaneously dock a series of ligands in unison instead of individually as single ligands. We hypothesize that this will increase the efficiency and accuracy of sampling. We illustrate RLEs hypothesized sampling advantage in Figure 2.1.

Due to the presence of functional groups of varying sizes found within a SAR series, there may be binding modes available to certain molecules but not others. RLE is capable of eliminating binding orientations not available to the ensemble as whole. Furthermore, highly similar ligands are expected to bind in similar fashions with common interactions to the chemical core[69, 2]. The RLE scoring function emphasizes favorable positioning for the common scaffold, shown by the red outline. The greater number of molecules that share a common substructure, the greater the scoring emphasis on that particular substructure. It is not anticipated that RLE will significantly improve docking for congeneric ligands that

Independent Single Ligand Docking
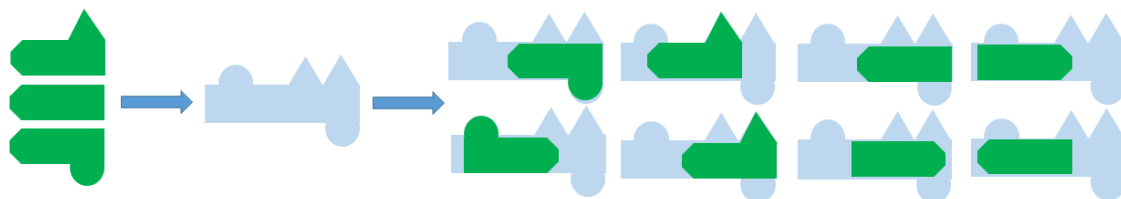
Simultaneous Ligand Ensemble Docking

Figure 2.1: Hypothesized mechanism of RLEs sampling advantage. Top: Three small molecules (green) are independently docked by RosettaLigand into the protein binding pocket (blue). There are multiple docked orientations possible for each small molecule. Bottom: The same three molecules are first aligned using their common scaffold (red). Docking in concert using RLE then yields a single, unambiguous binding orientation.

exhibit significantly different binding modes. Malhotra et. al. reviews receptor and ligand characteristics that tend to exhibit these alternate binding modes[70].

2.3   Experimental Methods

The validation dataset of 89 protein-ligand cocrystal structures curated across twenty systems is described in Appendix .

2.3.1   RosettaLigandEnsemble algorithm

Figure 2.2 illustrates the two-stage RLE algorithm. RLE takes as input a single protein structure and a congeneric series of molecules superimposed by chemical scaffold. In the low resolution TransformEnsemble phase, the same 3D translations and rotations are applied to all molecules in order to maintain the superposition and find a common binding mode. Step sizes and direction for both translation and rotation are taken from a Gaussian

distribution centered on a user provided value. Scoring is done using a pre-generated shape complementarity energy grid and moves are accepted/rejected by a Metropolis Monte Carlo criterion based on the sum of scores for all ligands in the ensemble. The protein structure remains static but ligand conformers are changed by swapping out individual ligands with alternate conformations from pre-generated libraries. The benchmark used the fragment based BCL::Conf small molecule conformer generator[95]. During the high resolution HighResEnsemble phase, only small perturbations to the ligand are applied with the focus on optimizing the protein-ligand interface. Since side-chain orientation differences are observed even for binding of related ligands, each protein-ligand interface is optimized independently. In a single simulation run, RLE generates x models where x is the number of ligands in the ensemble. Over the course of n simulation runs, RLE generates n*x total models, the same quantity as x independent RosettaLigand runs of n trajectories each.

The bulk of the computation time in both RosettaLigand and RLE is due to protein side-chain rotamer sampling during the high resolution docking phase. Since RLE generates individual protein-ligand models for the high resolution stage, computation time is not significantly altered.
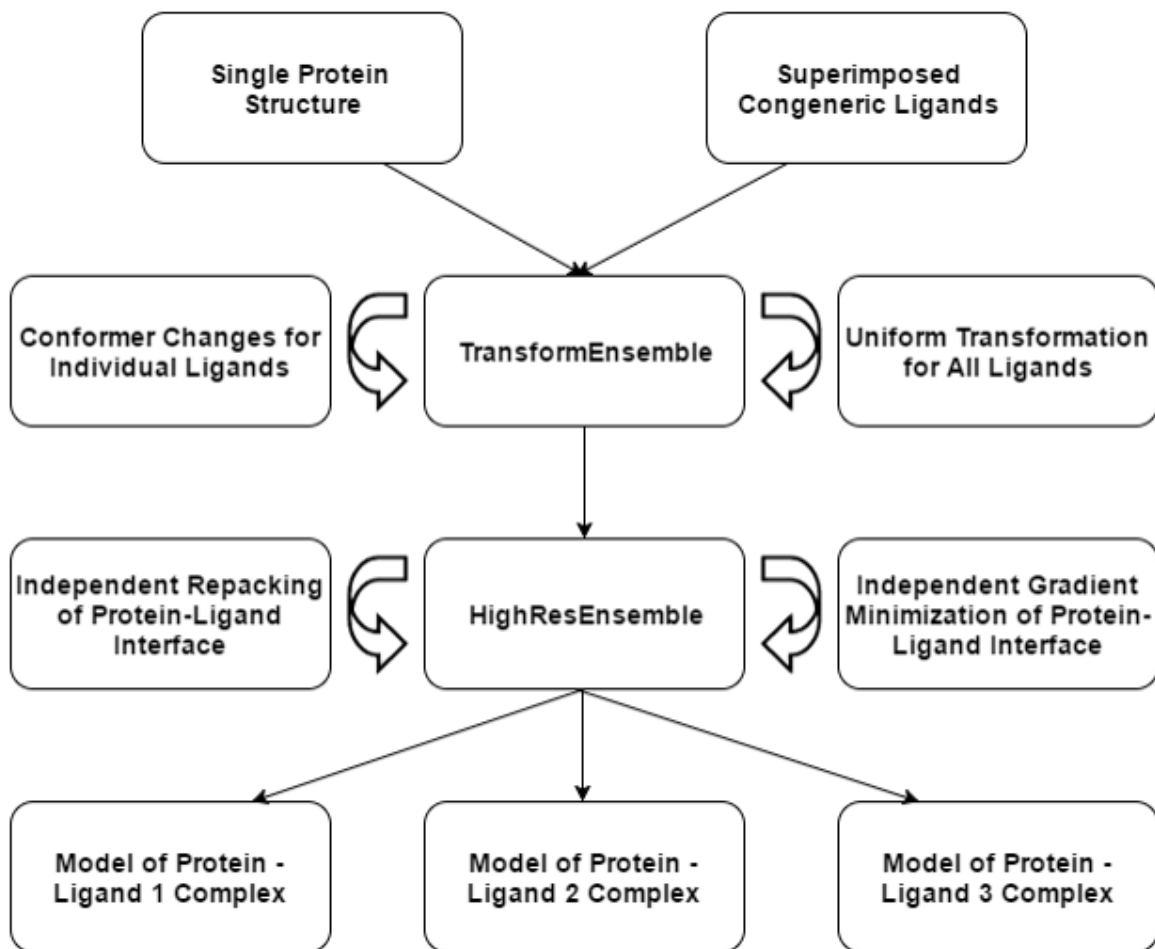
Figure 2.2: Illustration of RLE algorithm. The algorithm is separated into the low resolution TransformEnsemble step and the high resolution HighResEnsemble step. Curved arrows represent repeated moves accepted or rejected based on Metropolis Monte Carlo criterion. Individual model of each protein-ligand pair are outputted from a single protein structure and superimposed congeneric ligands as input.

2.3.2   Experimental model generation

Initial parameters for RLE are derived from the latest features of RosettaLigand algorithm[18, 20] and optimized for sampling efficiency. Additional sampling cycles and a decreased rotational barrier was necessary to counteract the increased sampling space involved in finding an optimal position for all molecules simultaneously. The exact number of sampling steps was calculated on-the-fly based on difference between the current step score and the maximum possible score assuming all atoms formed favorable interactions. Meanwhile,

the repulsive score term was halved to allow the entire ensemble to rotate through clashes. Ligand atoms are forbidden from moving outside of the defined docking sphere as was the case in RosettaLigand.

Following optimization, docking was performed with both RosettaLigand and RLE and evaluated for native ligand pose recovery. For each system, individual molecules were docked independently and as an ensemble into the same receptor structure. For each run, 2500 models were produced and the top ten percent were selected based on ligand interface energy for subsequent analysis.

In order to make the docking simulation resemble actual use, a uniform volume random translation within a 5 sphere and a random full rotational orientation is performed prior to docking. A random conformer is selected from the ligand conformer library. This avoids biasing the starting position and orientation to that observed in the crystallographic complex. An example of how to generate models for one system is provided in Appendix B.

2.4   Results and Discussion

We examine the top ten percent of scoring models by ligand interface score for each ligand cross-docking case. The top 250 models are analyzed for both sampling efficiency and scoring discrimination of native-like models. Native-like models are defined as having a ligand root mean squared deviation (RMSD) of less than 2 compared to the co-crystal structure. Sampling efficiency is represented as the percentage of models that are native-like, while scoring discrimination is represented as the scoring rank of the first native-like model. A higher sampling percentage of native-like models and a lower scoring rank for the best scored native-like model indicate improvement.

2.4.1 RLE improves sampling and scoring among top models

Among the top ten percent of models by score, RLE improved both the percentage of native-like models and the scoring rank of the first native-like model when compared to RosettaLigand. The increased sampling efficiency was observed in 62 out of 89 cases while the improved scoring rank was seen in 22 out of 89 cases as shown in Figure 2.3. In three cases, RLE produced a native like model while RosettaLigand did not. In ten cases, neither RLE nor RosettaLigand were able to find a native-like model in the top ten percent.
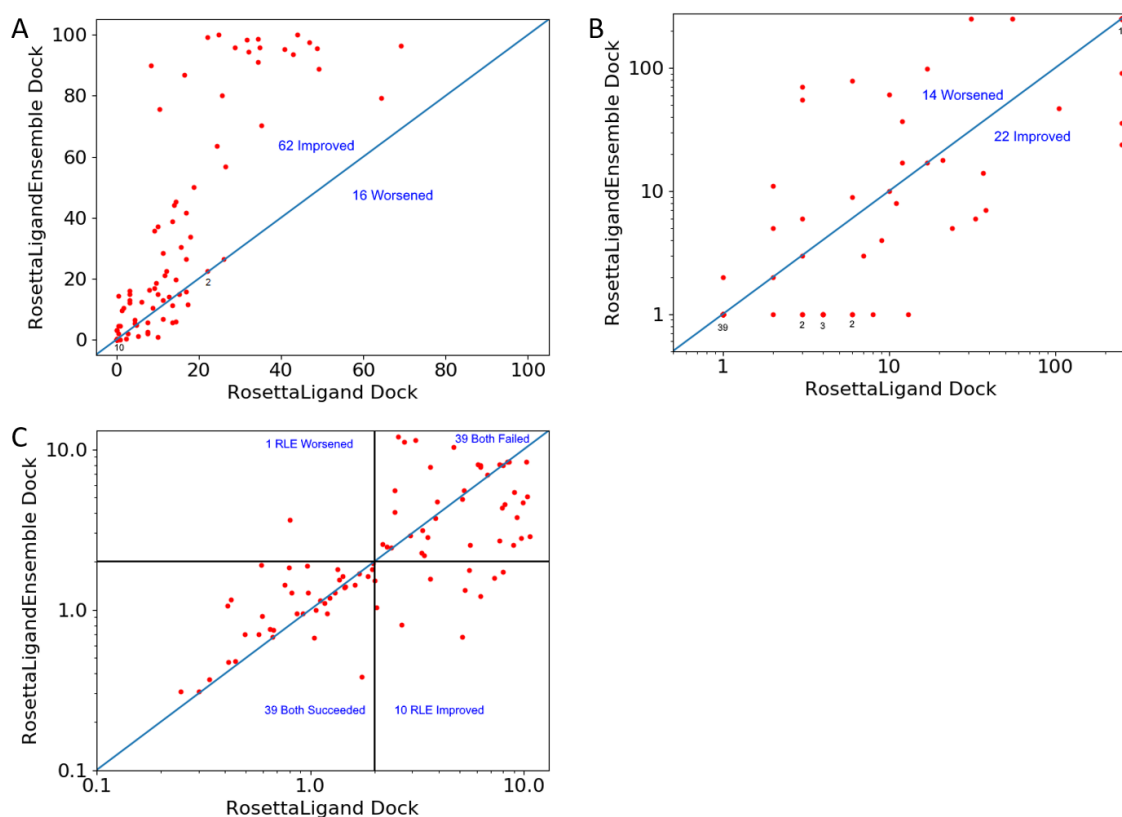
Figure 2.3: Comparison of sampling efficiency and scoring discrimination among top ten percent of models by score from individual RosettaLigand docking versus ensemble RLE docking. Overlapping dots are indicated by number of overlapped points below it. Blue diagonal line shows when RosettaLigand and RLE performance are identical. A: Percentage of native-like models from single and ensemble docking B: Scoring rank of the best scored native-like model from single and ensemble docking C: Small molecule RMSD of the top ranked model from single and ensemble docking. The 2.0 angstrom success cutoff is marked out in black lines.

Among cases where RLE improved sampling efficiency, nearly half saw an improvement of at least 25 percent. In contrast, no case saw RLE decrease sampling efficiency by more than 9 percent. For scoring discrimination, RLE recovered a native-like top scoring model in ten cases where RosettaLigand failed to do so. This is important as RLE would have still produced an accurate model in an application scenario even for these cases. There is a single case where only RosettaLigand produced a native-like top scoring model. Here, RLE still produced a native-like model in the top ten scoring.

Although the sampling efficiency increase was significant, there does not appear to be a direct translation between the number of native-like models and the ability to discriminate them from non-native-like models. Since both algorithms utilizes the same knowledge based scoring function during the high resolution docking and the final ranking, its expected that they may have a similar model discrimination power. This is illustrated in Figure 2.3c where in the large majority of cases, RLE and RosettaLigand either both succeeded or both failed at ranking a native-like model as the best scoring. However, there are ten cases where RLE was able to rescue the performance of RosettaLigand by producing a native-like best scoring model. Averaged across all 89 cases, the sampling efficiency improved by 18 percent and in 20 of these cases, both sampling and scoring metrics improved.

2.4.2 RLE eliminates alternate binding modes

The final binding location and orientation of the ligand is primarily determined by the low resolution docking stage. Perturbations of the ligand in the high resolution stage are minimal as the bulk of computational time is spent towards conformational energy minimization of protein sidechains. The RLE low resolution phase moves all molecules in unison, maintaining superimposition, and therefore force molecules to adopt a common binding mode. This coordinated movement is the process that eliminates binding volume available to some but not all members of the group. Figure 2.4 shows the ligand RMSD distributions seen among the top ten percent of scoring models for both RLE and Roset-

taLigand docking. Each protein-ligand pair of the system is plotted separately so effects across the system can be observed. Higher density at the low RMSD end of distribution indicates success. The red line in each subplot shows the 2 ligand RMSD cutoff for native-like binding modes. The systems have been sorted qualitatively into broad categories based on whether or not RLE generally improved both the sampling efficiency and the scoring discrimination. The RMSD distribution pattern for RLE is much more consistent within a system than the RosettaLigand distribution patterns for the same system. This is the aforementioned forced common binding mode effect. However, there remains individual protein-ligand pairs within a system where the distribution was not significantly improved.
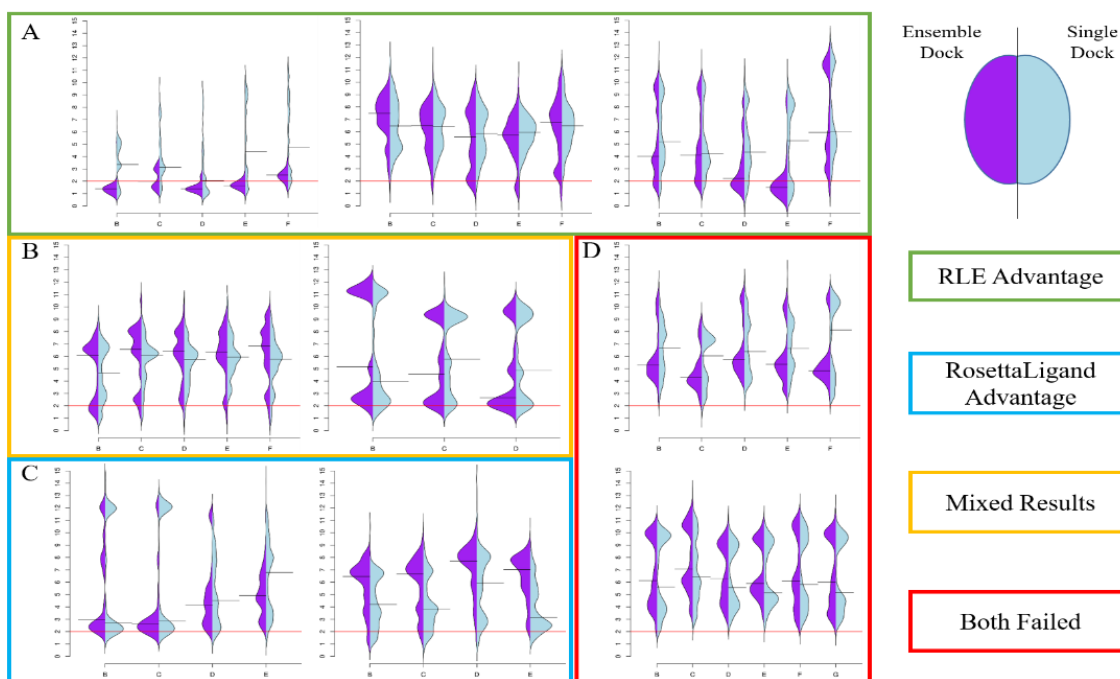
Figure 2.4: Ligand RMSD distribution observed among top 10 percent of models for RosettaLigand and RLE. Nine example systems have been separated into four qualitative categories of sampling and scoring change. For each system, data is split by individual protein-ligand pair with RLE docking on the left and RosettaLigand on the right. Black line shows the median and red line shows the 2 cutoff. A: RLE improved docking sampling and scoring for CTAP, HCV, and TPPHO (left, mid, right). B: RLE improvement varied from ligand to ligand within system for CDK2 and P38 (left, right). C: RosettaLigand performed better for LPXC and THROM (left, right). D: Both methods failed to perform well for CATB and THERM (top, bottom.

In the systems where RLE drives both a sampling efficiency and a scoring discrimi-

nation improvement (green), RLE eliminated a significant number of high RMSD binding modes seen in the RosettaLigand results. In the CTAP example, RLE ligand RMSDs are all within a similar range while the outliers produced by RosettaLigand are eliminated. It remains possible for ensemble docking to be more successful for certain ligands within group than others. Ligand C for CTAP has a smaller second peak that is not consistently eliminated by ensemble docking. One reason for this is because the high resolution stage considers ligand conformers in additional to protein conformers. For larger, more flexible molecules, RMSD may be relatively high even if the correct binding location and orientation is recovered. This is due to ligand conformational flexibility in the distal regions. Alternatively, in the HCV example, the majority of models from both RosettaLigand and RLE are not native like but only RLE generates a batch of native-like models. This is the aforementioned rescue scenario in which RLE is able to produce a correct model when RosettaLigand cannot.

A limitation to the RLE algorithm occurs when the alternate, high RMSD binding mode is available to all molecules within a system, as seen in the P38 system with mixed results (orange). RLE does not provide a significant advantage in scoring discrimination when both methods have a similar sampling efficiency. The emphasis on placement of the common scaffold means that an incorrectly identified common binding mode will result in poor performance across the system as seen in THROM (blue). RosettaLigand was able to produce good results for two members of this system because its docking runs are independent. Whether or not the different binding modes will be correctly consolidated in an actual application depends on the particular post-hoc analysis chosen. This incorrect placement of the common scaffold is repeated in the CATB and THERM systems where the distribution peaks fall out of the native-like RMSD range (red). One reason for this is the lack of chemical diversity in the functional group modifications within the group. These systems are difficult cases that neither algorithm can dock well.

2.4.3 Illustrative examples of success and failure

The binding pocket for several illustrative examples are shown in Figure 2.5. CTAP ligand B and TPPHO ligand C both show a significant improvement in sampling and scoring. The best scoring RLE model is native-like in both cases and sampling efficiency was 2.1x and 3.3x better for TPPHO and CTAP respectively.
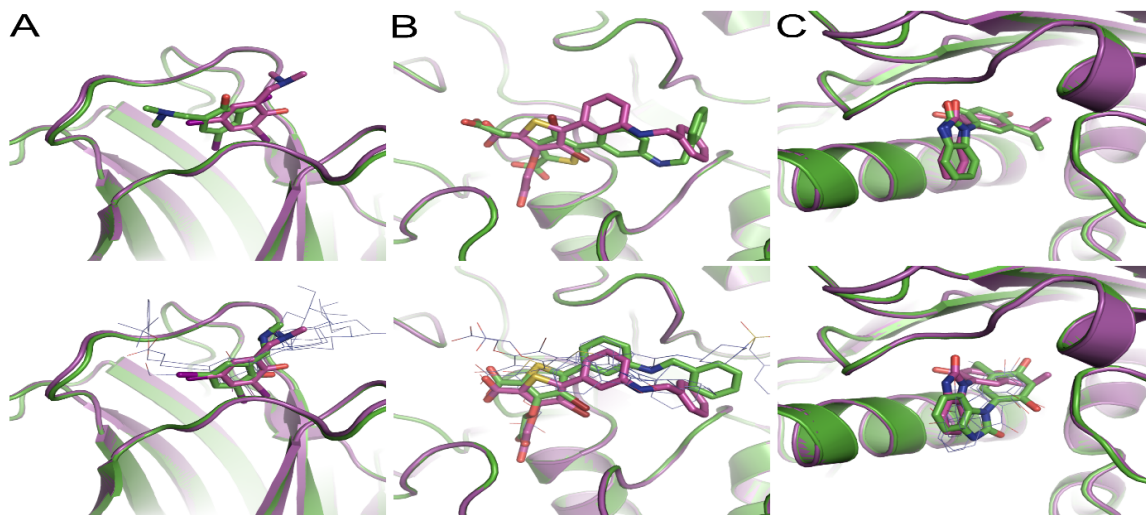
Figure 2.5: Illustrative examples of success and failure in recovering a native-like best scoring model. The top panels show the best scoring model from RosettaLigand and the bottom panels show the best scoring model from RLE. The co-crystal structure is shown (purple) aligned with the model (green). Remaining ligands of the RLE ensemble are shown as blue lines. A: CTAP system, ligand ID B (PDB: 4AGL) B: TPPHO system, ligand ID C (PDB: 2QBR) C: HSP90 system, ligand ID B (4YKQ).

In the CTAP example, ligand B is a small ligand in a relatively open binding pocket. This made it difficult for RosettaLigand to determine the proper orientation, generating three equal possibilities as shown in the RMSD distribution in Figure 2.4. However, the remaining ligands built off of the core scaffold have large chemical modifications off of two sites. The interactions formed by the distal groups with the bordering protein loops allows RLE to identify the proper orientation of the common core. Another example of this orientation flip is illustrated in the TPPHO example. Although the RMSD distribution for ligand C is more distributed, a clear binding mode is available for ligands D and E in the same system. This gives RLE a modest increase in successfully orienting ligand C, with

the major deviation due to conformation rather than orientation.
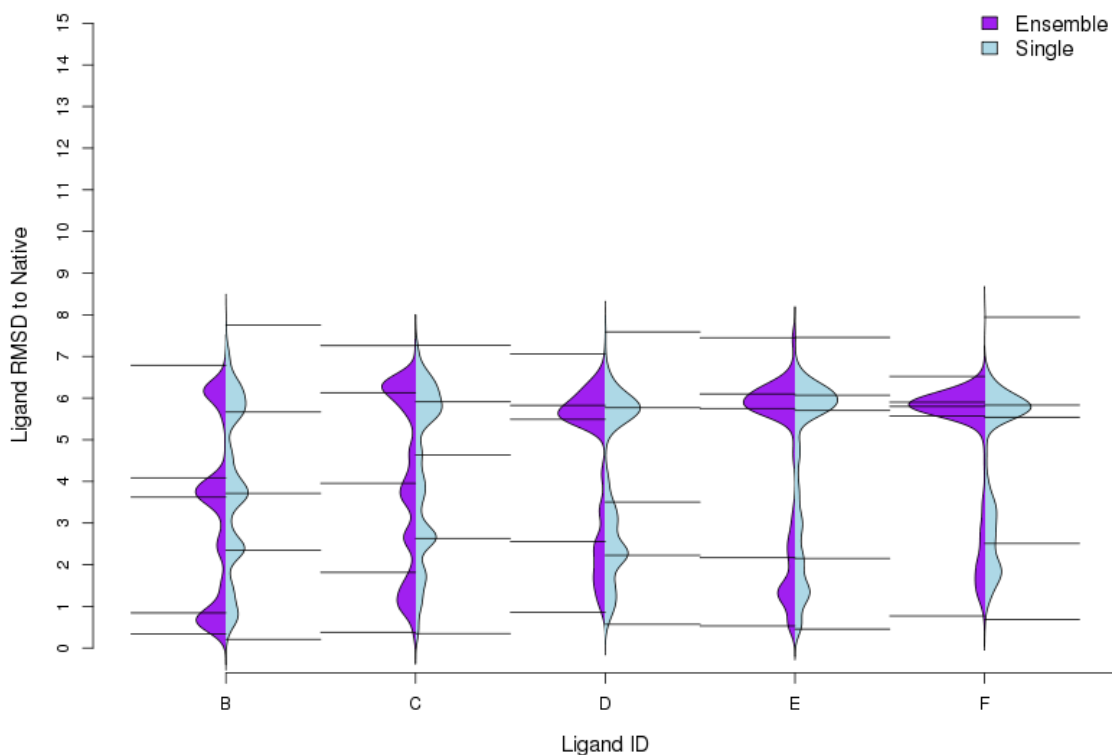
Figure 2.6: Ligand RMSD distribution observed among top 10 percent of HSP90 models for RosettaLigand and RLE. RLE docking distribution is on the left and RosettaLigand on the right. Five line summary shows highest, lowest, median, and quartiles.

Although RLE showed a slight sampling improvement for HSP90, the system proved to be a relatively challenging case for ensemble docking due to the small size of the ligand system. RosettaLigand produced a native-like best scoring pose while RLE generated a flipped conformation. The RMSD distribution however favors RLE with more native-like models. Across the system, there is a persistent alternate binding mode suggested by RLE, as seen in Figure 2.6, due to the fact that the binding pocket is much larger than the ligand. RLE is unable to rule out alternative binding modes of the common scaffold without distal groups that can eliminate conformational space.

2.4.4 Higher chemical similarity promotes higher sampling efficiency up to a limit

In order to better understand indicators of successful and unsuccessful systems, we sought to characterize the similarity of the cross-dock small molecules compared to the co-crystallized molecule. The traditional Tanimoto similarity coefficient is not particularly robust for more complex substitutions as it focuses on atom identity within a common substructure. We compared molecules using the in-house BioChemicalLibrary to calculate a PropertySimilarity[59]. PropertySimilarity measures similarity based on atomic charges, Van der Waals volume, bond types, and the presence of hydrogen bond donors/acceptors. For this dataset, PropertySimilarity has a general positive correlation with Tanimoto similarity as shown in Figure 2.7.
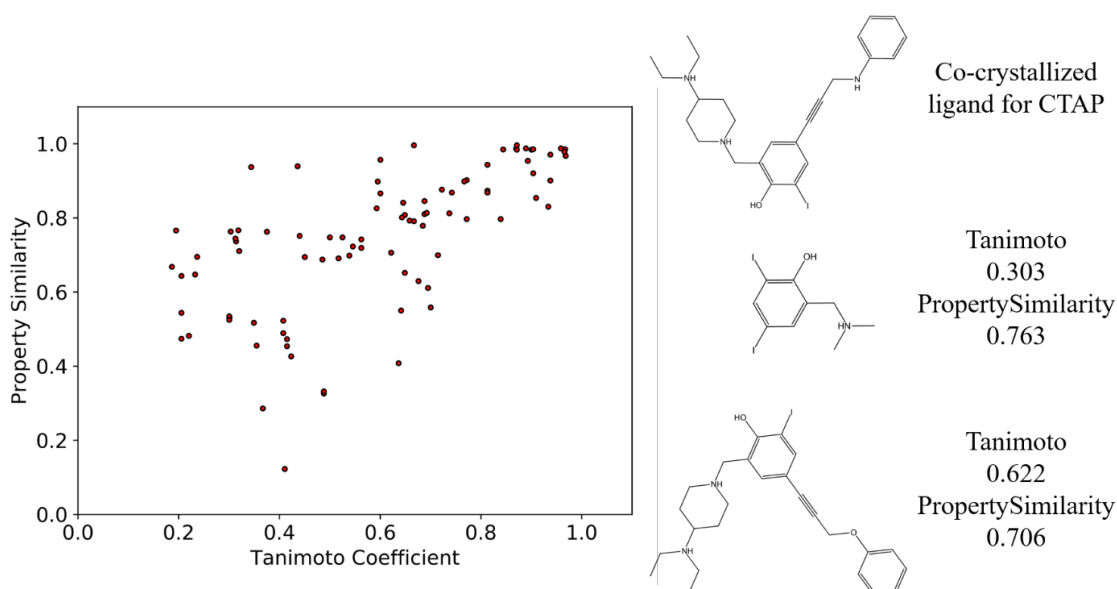
Figure 2.7: Tanimoto Similarity versus Property Similarity for 89 cross-docking test cases. Each molecule is compared to the ligand co-crystallized with the receptor structure for the system it belongs to.

The relationship between sampling and scoring improvement to property similarity is shown in Figure 2.8. Ligands are classified based on whether both sampling and scoring improved (red), both worsened (blue), or a mix of the two (white). There is a general tendency for molecules docked with high sampling efficiency to have a high chemical similarity to

the co-crystallized molecule. However, there are a number of highly related molecules in Figure 2.8 that are poorly sampled, suggesting that chemical similarity is a necessary but not sufficient condition of docking success. This is in agreement with previous studies that have shown docking success increases with chemical similarity[80, 96].
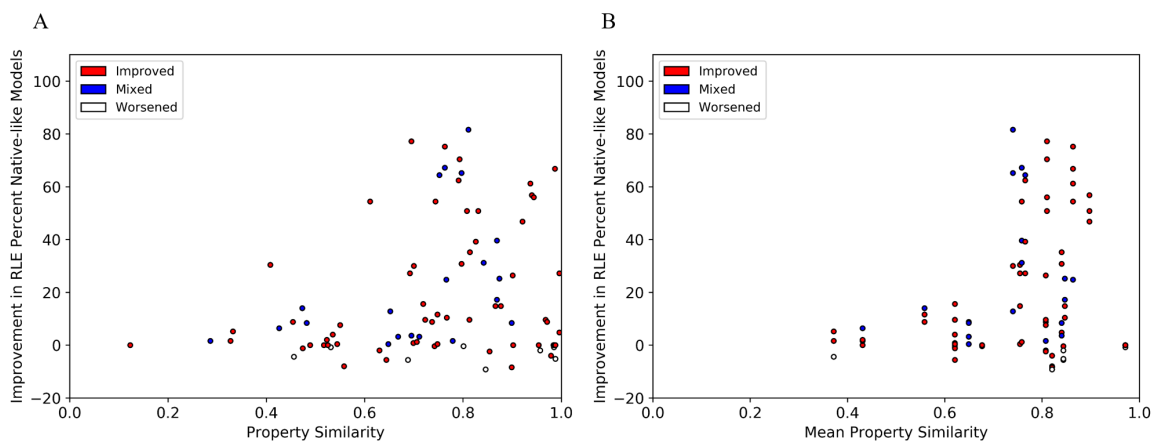
Figure 2.8: Sampling efficiency versus PropertySimilarity for top 10 percent scoring models. Protein-Ligand pairs are divided into cases where RLE improved both sampling and scoring (red), worsened both sampling and scoring (white), or improved one but not the other (blue). A: PropertySimilarity of each test ligand to the ligand co-crystallized with receptor structure vs the improvement in RLE docking as calculated by RLE percent native-like minus RosettaLigand percent native-like. B: The mean PropertySimilarity for each protein system vs. the improvement in RLE docking. Each vertical set of dots represents a single protein-ligand system.

In order to predict performance on a system level, we computed the mean PropertySimilarity for each system and plotted this value against each ligands improvement in sampling efficiency when docked with RLE. This is shown in Figure 2.8 with each vertical line comprising of a congeneric set of protein-ligand pairs.

The largest improvement falls around a mean PropertySimilarity measure of 0.8, suggesting that there is a sweet spot for improvement. Systems that are too different (P38, mean=0.37) or too similar (THERM, mean=0.97) exhibit limited benefits from ensemble docking. In particular, the THERM system consists of a chemical scaffold to which the primary modification is the switching of various hydrocarbon groups. Furthermore, the molecule interacted with two separate hydrophobic ends and a network of water molecules

in the binding account, which makes orientation determination difficult[98].

2.4.5 Identifying favorable binding poses corresponding with SAR data

The inaccuracy of ranking despite accurate docking remains problematic. One post-hoc solution is to select sets of binding modes that correlated with experimental data. We sought to address this deficiency during docking by adding a corrective factor to drive high resolution docking towards binding modes. Following each cycle of optimization, we modified the scoring difference based on the Spearman correlation to experimental data as an adjustment prior to applying the Metropolis criterion. The adjustment provides an additional bonus to perturbations that improved the score of stronger binding or more active ligands, and to perturbations that worsened the score of weaker binding or less active ligands. The low resolution docking stage remains the same and does not account for experimental correlations. The adjusted co-dependent algorithm is shown in Figure 2.9a with the score adjustment being applied in the highlighted step.

Although the corrective factor does improve correlation with experimental affinity, it does not improve the sampling efficiency or docking accuracy. This is in part due to the fact that the binding orientation is primarily determined in the low resolution phase that does not account for correlation. The Spearman correlation coefficient is defined as the Pearson correlation based on only the ranks of the models. Therefore, the Spearman correlation has a discrete distribution with limited values available for a small data set. This makes it difficult to significantly improve the correlation in many cases. The results are in agreement with previous results showing that improvements in the Pearson scoring correlation using machine learning based scoring functions only translated to a moderate increase in accurate ranking[99].

One additional hindrance to a more successful corrective method is the dilemma in selecting models illustrated in Figure 2.9b. In order to maintain the experimental correlation, entire ensembles of ligand models must be selected. However, since the Monte Carlo
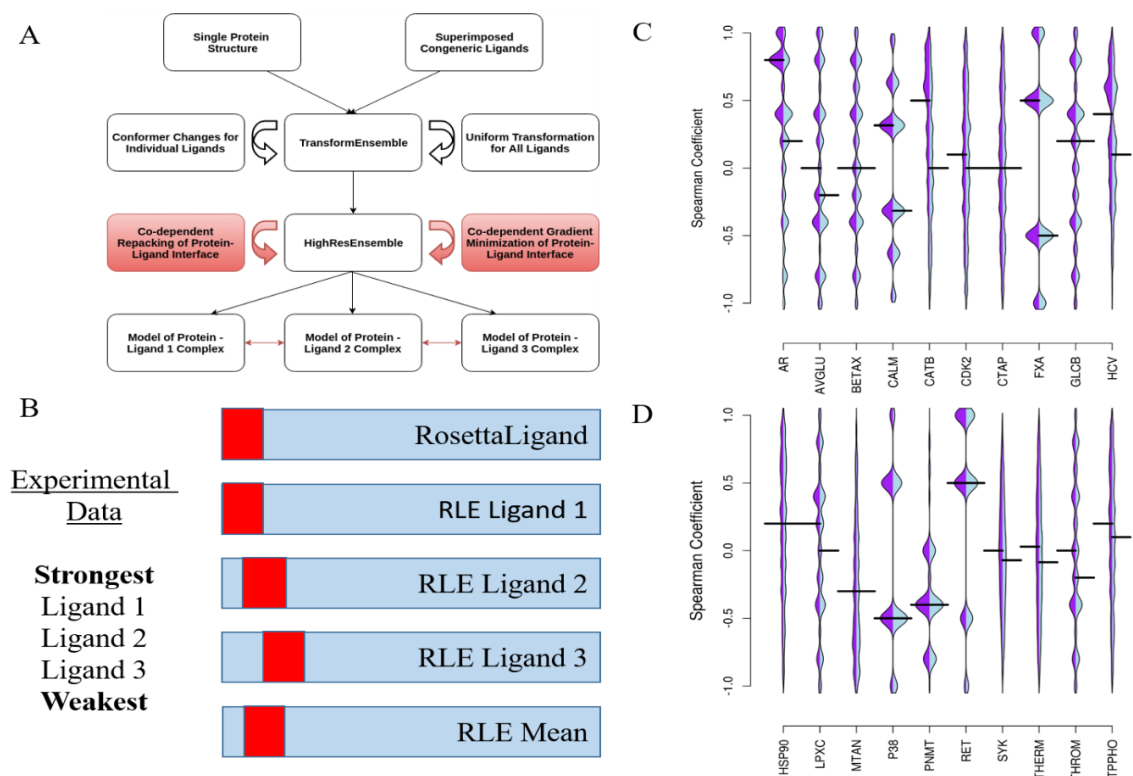
Figure 2.9: RLE Spearman correction during high resolution docking to favor binding modes that correlate with experimental data. A: Spearman corrected RLE high resolution docking steps shown in red B: Model selection dilemma resulting from inaccuracies in docking scoring function C,D: Distribution of Spearman correlation in generated ensembles for each system with a corrective factor (left, purple) and without (right, blue). The mean line is shown in black.

sampling method is stochastic, it is unlikely that each ensemble will contain low energy conformations of every protein-ligand interface. Selecting models by a mean metric across the entire ensemble may select the best scoring models for one ligand but not for others. Even with improvements in scoring functions, this selection dilemma may prevent RLE from simultaneously selecting the best models for each ligand.

2.4.6 Comparing RosettaLigandEnsemble with protein-ligand docking tools

We used AutoDock[9] with a Lamarckian Genetic Algorithm to test its performance in the 89 cross-docking systems in the benchmark set. Standard protocol settings and

a docking volume comparable to RLE docking were used to generate the models. The AutoDock simulations were performed using a rigid receptor model.
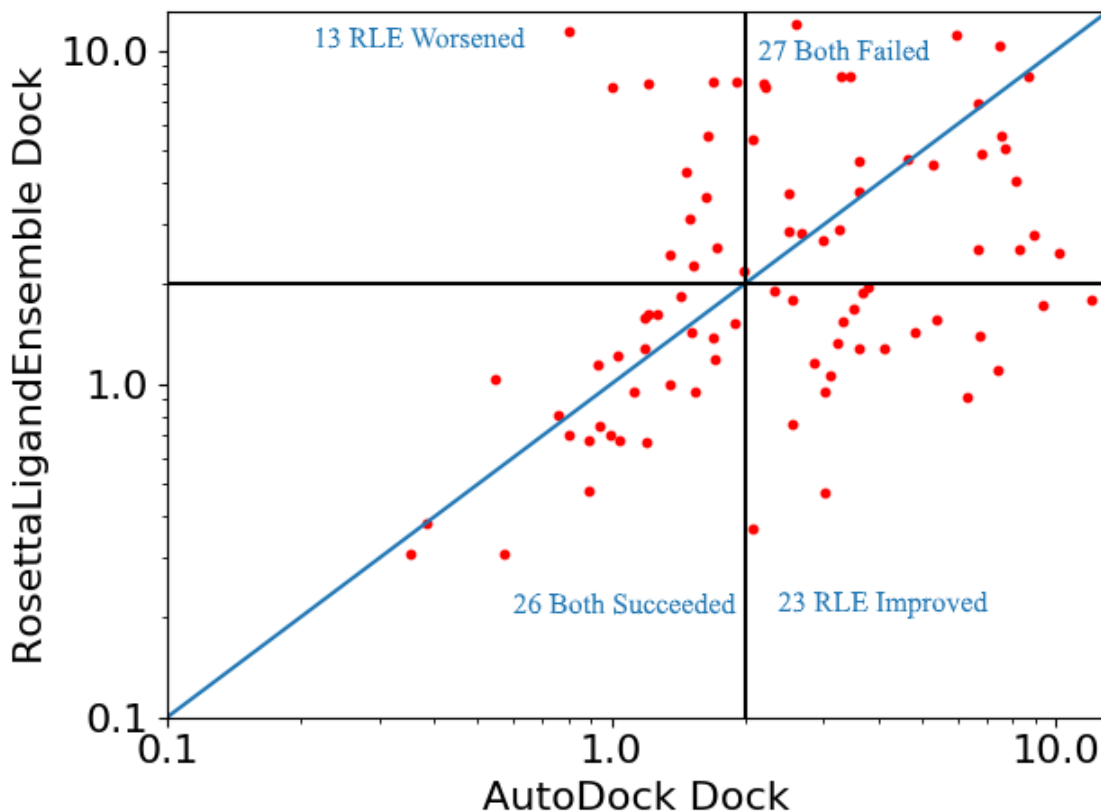
Figure 2.10: Small molecule RMSD of the top ranked model from RLE and AutoDock docking. The 2.0 angstrom success cutoff is marked out in black lines. Equal performance of the two software is indicated by the blue line.

Figure 2.10 shows the small molecule RMSD of the top scoring model from RLE and AutoDock docking. RLE recovered a native-like small molecule pose in 23 cases where AutoDock did not. By contrast, there were only 13 cases where AutoDock had a native-like best scoring model when RLE did not.

Wang et. al. evaluated ten docking software across the PDBBind dataset, including 18 cross-docking cases from the present benchmark in which at least one tested method did not recover a native-like top scoring model[100]. RLE rescued the performance in 15 out of 18 of the cases in Figure 2.11.

System	PDB	RLE	AutoDock (LGA)	AutoDock (PSO)	AutoDock Vina	Le Dock	r Dock	UCSF DOCK	Ligand Fit	Glide (SP)	Glide (XP)	Gold	MOE Dock	Surflex Dock
AVGLU	4IO3	4.12										3.85		
AVGLU	4IO7	2.32						4.59						
BETAX	1FHD	6.64			2.27			3.24	6.57	6.87		5.56		5.58
CALM	3SXF	0.49	6.44		6.45									
CALM	3T3U	1.04	5.32		5.34									
CALM	3V5P	1.13	6.45		4.24									
CALM	3V5T	1.71	5.41		5.79									
MTAN	1JYS	0.57	2.49	2.50		2.67				2.89			2.60	
MTAN	1NC1	0.55	4.85		4.79									
MTAN	1NC3	0.94	2.39	2.41	2.48									
MTAN	1Y6R	0.99	4.80		5.04									
P38	2ZB1	2.07	8.69		3.98			10.70						
PNMT	1HNN	0.93								5.25				
PNMT	2G70	1.03								2.09				
THROM	2ZFP	7.39						5.32						
TPPHO	2B07	2.67	2.53		2.61		2.38	2.06		2.19		2.27	2.45	2.10
TPPHO	2QBR	5.37						2.27						
TPPHO	2ZN7	1.20						2.21		3.80				

Figure 2.11: RosettaLigandEnsemble performance on 18 failure cases from Wang et. al. The table shows small molecule RMSD of the top ranked models from RLE and ten other docking software. All results except for RLE were pulled from Wang et. al.

Most notably, RLE was able to generate native-like models across the calcium-dependent protein kinase CDPK1 (CALM) and Helicobacter pylori nucleosidase (MTAN) systems. However, the 2B07 test case from a series of protein tyrosine phosphatase inhibitors (TP-PHO) remains challenging. This is likely related to the orientation flip discussed with regards to Figure 2.5. RLE was also able to recover native like top scoring models for all five spleen tyrosine kinase (SYK) compounds also tested in CSAR 2014, matching the performance of the best available docking tools.[101] However, RLE performed worse on an deacetylase (LPXC) test system that was part of CSAR 2012[24], only recovering a near native model in 1 out of 4 cases. The generated best scoring models had a RMSD of 2.23 in the worst test case, suggesting only a minor performance decrease in the LPXC system. It should be noted that these comparisons do not account for additional protein flexibilities accounted for by RLE, nor does it include the effects of differences in starting ligand conformation. However, there does not appear to be a strong induced fit or conformational selection component in these structures.

2.5 Conclusions and Future Directions

2.5.1 Needed improvements in decoy discrimination

The improved sampling efficiency did not directly translate into improved scoring rank-ing partly due to the inaccuracies in discriminating between native-like and non native-like models. Better decoy discrimination in conjunction with the more efficient sampling will allow for fewer models to be produced before converging on a native like binding mode. The reduced number of models will greatly reduce the time and computational resources necessary for docking. Furthermore, the SAR correlated docking would benefit greatly from a more accurate scoring function capable of ranking ligands. RLE in combination with such a method would generate binding modes in accordance with SAR data without the need for post-hoc filtering.

2.5.2 Consideration of alternate binding modes among congeneric ligands

RLE docking is generally designed for docking in cases where similar ligands exhibit a common binding mode. This is the case for the vast majority of known protein-ligand crystallographic complexes[2, 70]. Presently, a priori assumptions are made for a given system, even if single ligand docking is used as initial placement is often based on previ-ously seen binding modes. A future development of RLE docking would allow for minor shifts in the binding mode while maintaining the general placement and orientation, a sort of soft ensemble docking. Furthermore, the use of a property based alignment method such as PropertySimilarity will allow for common scaffolds based on chemical similarity as op-posed to identity. Cases wherein similar ligands bind in completely different pockets or to different protein conformations will remain challenging for ensemble based methods.

2.5.3 Ensemble approaches from protein structure based direction

A similar approach can be used to drive ensemble docking improvements in the use of protein mutation data. Current approaches to protein ensembles generally focus on accounting for conformational diversity. Mutational data on proteins is used to identify potential protein-ligand interaction sites as a distance restraint to docking. An alternate ensemble approach would utilize SARs based on multiple protein mutants to determine how the ligands may bind to each mutant within the series. A further step would be in combining protein ensemble and ligand ensemble methods to improve docking accuracy by considering how ligand modifications fit into the different pockets of protein mutants. Multi-target virtual screening, in particular with biologically relevant mutants, can be performed with such an algorithm. The Platinum database of small molecule interactions with protein mutants[102] provides an excellent source of data for training an algorithm in this approach.

Chapter 3

ROSIE Ligand Docking

3.1 Summary

This chapter discusses the upgrade of protein-small molecule docking server for use by non-computational focused researchers interested in modeling a protein-ligand interface. The server was updated to use the improved RosettaLigand docking algorithm described in DeLuca et. al. [18]. Further additions such as the inclusion of a small molecule conformer generator were made with the goal of reducing manual preparation steps necessary for a docking submission.

The work in this chapter is ongoing but the bulk portion will be submitted as Fu et. al. "ROSIE Ligand Docking: A Protein-Small Molecule Docking Server" for which I am an equally contributing first author.

3.2 Introduction

Many proteins function by interaction with endogenous small molecule ligands. Small molecule therapeutics account also for the majority number of FDA-approved medications[103]. Therefore, understanding the function of proteins as well as developing of novel drug candidates requires a structural understanding of binding interactions between a protein and a small molecule. Atomic level insights regarding these interactions from structures of protein-ligand complexes determined by X-ray crystallography are ideal for this purpose, but experimental structures can be challenging and costly to obtain. Computational protein-ligand docking aims to provide a full-atom model of the bound protein-small molecule complex using existing structures of the individual partners as the starting point. Recent advances and applications in small molecule docking and drug discovery have been dis-

cussed extensively[104, 87].

A number of protein-ligand docking servers are presently available. SwissDock[105] is a server based on EADock DSS, an evolutionary algorithm, while MEDock[106] uses a novel optimization algorithm that works well with more rugged energy landscapes. PatchDock[107] is a geometry based method focused on shape complementarity. There are also standalone docking tools such as NRGsuite, which integrates with PyMOL to allow users to setup real-time simulations on their own machines[108]. While some structure-based virtual screening servers can also perform docking, these tend to focus on speed for docking large libraries rather than accuracy for recapitulating a single protein-ligand interface. Other small molecule docking methods generally require the user to provide computational resources (e.g. AutoDock[9]) or a fee for usage (e.g. DockingServer[109]).

RosettaLigand[16, 17] is a well-established small molecule docking protocol within the Rosetta biochemical modeling software suite[1, 110]. RosettaLigand uses shape complementarity to sample reasonable binding modes and then performs high resolution refinement using a knowledge based scoring function. The algorithm treats the ligand and surrounding protein residues, both backbone and side-chain, as fully flexible during the refinement process. Hundreds of independent simulations can be quickly generated by RosettaLigand running on a cluster.

There are a few limitations of the existing implementation of the RosettaLigand application: 1) Rosetta primarily utilizes a command line interface, 2) Rosetta has specific input formats for protocol scripts and ligand parameter files, 3) Rosetta has a large variety of options for customizing the docking process, and 4) analysis of outputs often require additional scripting. While the ability to customize protocols is an advantage to the experienced user, it may be overwhelming for newcomers to protein-ligand docking. In order to overcome these challenges, we have developed the ligand docking protocol for the Rosetta Online Server that Includes Everyone (ROSIE)[111].

3.3 Experimental Methods

ROSIE Ligand Docking utilizes a two stage approach for predicting the binding mode between a small molecule ligand and the protein target. A low resolution first stage identifies likely binding modes based on geometric complementarity of the ligand with the protein binding pocket. A rigid protein model is used to generate an attraction/repulsion grid based on hard spheres. The ligand is then allowed to translate, rotate, and change conformations in the docking volume with moves accepted or rejected based on a Monte Carlo Metropolis (MCM) criterion[18]. By default, the grid is 15 wide but users may increase the search space up to 30 .

The preliminary model from the low resolution stage is passed on to a high resolution MCM stage with a full-atom knowledge-based scoring function. Small perturbations are made to the ligand position followed by rotamer library sampling of local protein residues. Local residues are dynamically defined as residues within 7 of the ligand in addition to sequentially adjacent residues to allow for backbone flexibility. The restriction of protein flexibility to interface residues makes modeling of large proteins tractable. A soft-repulsive energy minimization allows residues to sample conformations beyond the rotamer library. Following six cycles of MCM docking, a final hard-repulsive gradient minimization is used to optimize ligand and receptor torsion angles, both backbone and sidechain. The score difference between the docked protein-ligand complex and the separated protein-ligand pair is reported in Rosetta Energy Units as the interface_delta score, an emulation of binding free energy change. The models with the top ten most negative interface_delta scores are presented as the best models[16].

The detailed terms of the high resolution scoring function have been covered elsewhere[112, 17, 113]. Key terms included in the ROSIE Ligand Docking score function are 1) Lennard-Jones attraction/repulsion potentials, 2) Lazardis-Karplus implicit solvation, 3) hydrogen bonding potential, and 4) empirically derived rotamer and phi-psi angle probability scores. In addition, the ligand docking specific scoring function includes the above terms and elec-

trostatic terms for the ligand with adjusted weights benchmarked on 100 native protein-ligand complexes[17].

3.4   Results and Discussion

The following describes the ROSIE Ligand Docking server setup including a description of the input and output process. Previous validation of the underlying RosettaLigand algorithm is also discussed.

3.4.1   Inputs for ROSIE ligand docking

A structure of the protein target in PDB format is required. This structure can be derived from experimental or computational methods. The ligand molecule should be provided as a single SDF-formatted file. The user may choose to supply within this file all ligand conformers to be considered during docking, or select for ROSIE to generate conformers using the BioChemicalLibrarys knowledge based rotamer library sampling algorithm[95]. ROSIE Ligand Docking does not perform binding site detection and thus an approximate starting location for the ligand is necessary. The starting location can be provided by setting the ligand in the binding pocket prior to uploading the SDF file and checking the Use the starting coordinates in the SDF option. Alternatively, the user can provide the starting point as X, Y, Z coordinates. In this case, the server will automatically translate the centroid, calculated by averaging the coordinates of all non-hydrogen ligand atoms, to the provided coordinates. The automated conformer generation and use of SDF coordinates were not available in previous implementations of the server. For more information on X, Y, Z coordinates in the PDB format, consult the HETATM description on page 190 of the PDB format guide (ftp://ftp.wwpdb.org/pub/pdb/doc/format_descriptions/Format_v33_Letter.pdf). Additional information on the SDF file, an extension of the MDL Molfile format, can be found at https://en.wikipedia.org/wiki/Chemical_table_file. In particular, the first block provides the coordinates used by the server to define the ligand position.

In cases where the approximate binding site is unknown, SiteHound-web[114] can be used to identify potential ligand binding pockets. The center coordinates output by SiteHound-web can directly be used as the starting X, Y, Z coordinates for ROSIE Ligand Docking.

Figure 3.1: Screenshot of input screen for submitting a new ROSIE Ligand Docking job

Default sampling parameters that work well for most cases are provided, but may be adjusted under Advanced Settings to modify search radius, step size, or step count. Users do not have to modify any parameters under Advanced Settings. For a larger binding pocket, users may wish to perform more extensive sampling by increasing the Maximum

radius to search, Size of the low-resolution grid, Size of the low-resolution translation, Number of Monte Carlo sampling steps, and Randomize the initial position parameters. Conversely, if the likely binding location is better defined through experimental data, the user may wish to decrease these parameters for a limited refinement docking. There are further options for customizing the number of high resolution protein backbone and side-chain optimization cycles. Optional parameters allow the user to specify job title, job description, and email address for notifications. Users should consult the documentation for additional details and usage examples. Figure 3.1 shows an example input screen for docking eticlopride to human dopamine receptor 3 (PDB: 3PBL).

3.4.2 Outputs for ROSIE ligand docking

Figure 3.2 illustrates the typical output from a single run of ROSIE Ligand Docking. The output is displayed in the Results section in the form of images and file links for the top ten best scoring models. An interface score versus total score plot is available as a quick visualization of scored models. A score table is provided in the standard Rosetta score file formatting, where each row represents a single, independent model and each column represents a score term.

The primary scoring term used to evaluate ligand docking models is interface_delta, which is located in the second column. This value is calculated in Rosetta Energy Units and is meant to emulate the binding energy of the ligand, with more negative values being better. In contrast, the total score evaluates the protein-ligand complex in its entirety, including the unmodified protein residues outside of 7 protein-ligand interface region. If the user wishes to sort by alternative scoring terms, the user may simply click on the column header for the desired scoring term. The server enables downloading of all input and output as a single TAR archive or individually through custom web-links. Summaries are also provided in JSON format for use with database processing. The job submission parameters used to generate results are also available.

Results

Ten lowest-interface_delta-energy structures created in the ligand-docking run:
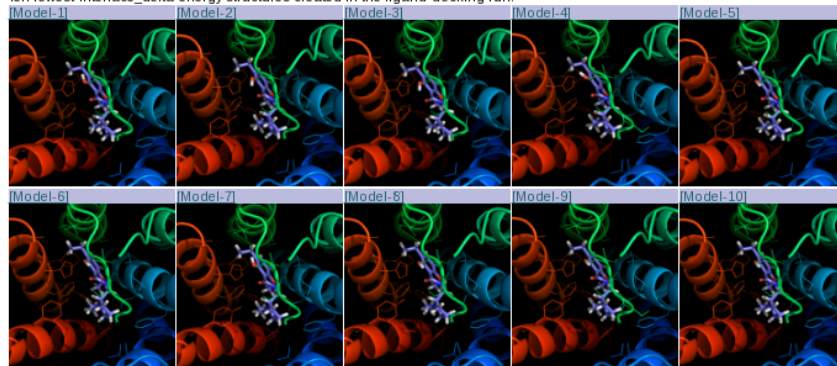
Interface Score/Total Score
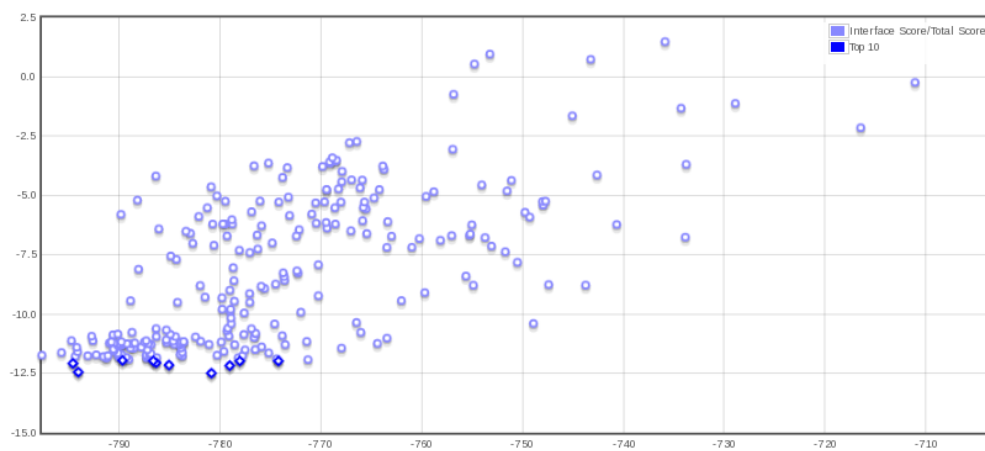
Hover over the graph points to see exact score and download particular result file.

Score data [Download original score file]									
decoy	interface_delta	total_score	Transform_acce	angle_constrain	atom_pair_cons	chainbreak	coordinate_cons	dihedral_constr.	dslf_ca_dih	dslf_cs_ar
protein_LG_0123	-12.495	-780.842	0.718	0	0	0	7.723	0	0.324	1.971
protein_LG_0147	-12.45	-794.041	0.788	0	0	0	7.977	0.299	0.321	2.059
protein_LG_0197	-12.175	-779.027	0.678	0	0	0	7.645	0.11	0.323	1.975
protein_LG_0048	-12.152	-785.05	0.754	0	0	0	7.145	0	0.323	1.955
protein_LG_0180	-12.082	-794.572	0.778	0	0	0	6.773	0.022	0.324	2.117

Figure 3.2: ROSIE Ligand Docking sample results page generated by docking eticlopride to human dopamine receptor 3. The output page includes the top ten scoring models generated, a score plot, a sortable score table, and download links to all files.

3.4.3   Information about ROSIE server

ROSIE Ligand Docking has completed more than 2000 ligand docking jobs since January 2014, which amounts to over 90,000 CPU hours. Roughly 1 CPU hour is needed to generate 100 structures with exact time dependent on ligand size and sampling parameters. Each docking model is an independent simulation allowing ROSIE Ligand Docking to generate all models on the order of real time minutes. The current ROSIE computing cluster is shared among all protocols and hence completion time from submission depends on present server load. Previous implementation of the ROSIE Ligand Docking utilizes the docking protocol described in Meiler & Baker (2006)[17] and Davis & Baker (2009)[16]. The present version combines the low resolution docking protocol from DeLuca (2015)[18] with the high resolution docking protocol previously used. The updated low resolution docking algorithm demonstrated a 10-15 percent increase in docking accuracy and a 30-fold speed increase over the previous method. ROSIE Ligand Docking has a Python front-end for user interface and records docking tasks into a MySQL database following input validation. Unseen to the user, the back-end converts user requests into Rosetta command-line scripts for the computer cluster. The results are stored back into the database for easy user access. This setup allows for developers to modify the modeling protocol without affecting user interactions[115]. ROSIE Ligand Docking is maintained and tested alongside the Rosetta Software Suite to ensure underlying protocol changes do not impact the server in unexpected ways.

3.4.4   Validation of RosettaLigand algorithm

The RosettaLigand protocol that serve as the basis for ROSIE Ligand Docking has been validated in a number of independent studies. Davis et. al. found the best scoring model to be within 2 of the native structure in 54/85 benchmark cases[16]. In a separate test on a pharmacologically relevant GlaxoSmithKline dataset, RosettaLigand obtained ¿40

percent docking success rate for 4 out of 7 systems. The current docking protocol was further refined in DeLuca et. al., which demonstrated an effective 30-fold speed increase and a 15 percent accuracy increase against 43 protein-ligand complexes[18]. These results were comparable to other docking software, all of which showed system-dependent success rates[16]. Absolute binding energy prediction on an HIV-1 protease/inhibitor dataset obtained a correlation coefficient R of 0.71[20].

The RosettaLigand algorithm can also utilize protein targets generated via comparative modeling. This is important as many pharmaceutically relevant targets do not have experimentally determined structures. Docking into comparative models is a more challenging task but success can be significantly improved by utilizing holo protein templates co-crystallized with chemically similar ligands. Kaufmann et. al. used RosettaLigand to recover a native-like binding mode in the top ten scoring for 21 out of 30 benchmark cases[36]. Nguyen et. al. demonstrated the applicability of this concept to G-Protein Coupled Receptors (GPCRs), a major protein family for drug discovery applications[59]. A broadly applicable comparative modeling method is not yet available on ROSIE but interested users are encouraged to look into the RosettaCM algorithm[116] or the combined Rosetta homology modeling plus docking protocol[117].

3.5   Conclusions and Future Directions

We have developed ROSIE Ligand Docking as part of the ROSIE set of easy to use, automated protocols for computational structural biology. Other available ROSIE protocols include protein-protein docking, peptide docking, and antibody modeling. Additional protocols can be developed via the directions provided in [115] and will be incorporated into ROSIE once server back-ends become available.

ROSIE Ligand Docking provides a lower barrier to Rosetta for users less familiar with scripting languages or computing clusters. Basic inputs of protein and small molecule structures is sufficient to generate models of the interaction complex. Furthermore, users

do not need to pre-compute conformations of protein or small molecules. More advanced users may customize docking options or use ROSIE Ligand Docking in conjunction with other computational methods. The standardized file formats of PDB and SDF for protein and small molecule respectively can used with other structural biology software with ease.

In the future, we anticipate improvements in the ligand docking protocol are necessary to keep up with drug discovery efforts geared towards more challenging targets. In particular, G-Protein Coupled Receptors and other membrane proteins remain difficult targets for computational ligand docking. The difficulty observed in scoring these membrane protein-ligand complexes may be improved by a better scoring function parameterized specifically for these ligands. The current ROSIE Ligand Docking implementation do not allow for customization of the scoring function. An iterative comparative modeling and ligand docking pipeline is also in development. Furthermore, ROSIE Ligand Docking does not currently allow for experimental restraints, a feature available in the command line version of RosettaLigand. The inclusion of protein-ligand interface restraints may help ROSIE Ligand Docking better identify binding modes that correspond to experimental data. The current workflow may also be upgraded to allow for use in virtual screening application.

Chapter 4

Applications of RosettaLigand and RosettaLigandEnsemble

4.1   Summary

This chapter discusses the application of protein-ligand structure prediction with Rosetta to small molecule discovery applications of pharmacological importance. Beyond RosettaLigandEnsemble docking, complementary techniques such as comparative modeling and loop closure are utilized to model protein receptors. The computational predictions in this chapter are checked in collaboration by research groups at the Vanderbilt Center for Neuroscience Drug Discovery and at Leipzig University.

The STAT ligands research discussed in this chapter contains material published as Lis et. al. "Development of Erasin: A Chromone-Based STAT3 Inhibitor Which Induces Apoptosis in Erlotinib-Resistant Lung Cancer Cells."[118] for which I am a middle author, and material from an additional co-author manuscript currently under review. The work on mGluR modulators is unpublished but contains a graphic reprinted with permission from Wenthur, C., & Morrison, R. (2013), "Discovery of (R)-(2-fluoro-4-((-4-methoxyphenyl) ethynyl) phenyl)(3-hydroxypiperidin-1-yl) methanone (ML337), an mGlu3 selective and CNS penetrant negative allosteric modulator", Journal of Medicinal Chemistry, 1(56), pp. 52085212, Copyright 2013, American Chemical Society[119]. Finally, the PAR binders research covers ongoing unpublished work.

4.2 Introduction

4.2.1 The necessity of comparative models

Research questions in pharmacology often focus on understanding how a particular small molecule ligand binds to a given protein target. With advances in cell-based assays for high throughput screening, the number of known protein-ligand interactions are being uncovered at a rapid pace[120]. However, Structure-based drug design requires detailed knowledge of molecular interactions beyond a simple measure of binding affinity. Experimental structure determination methods are improving rapidly and nearly 70 percent of the human proteome either have a known structure or have a homolog (>30% sequence identity) with a known structure[121]. In particular, in a structural coverage study of 667 human proteins targeted by 1194 approved drugs, roughly half had a determined structure and almost all had a homolog with a determined structure[122]. Comparative modeling is the computational technique for generating a protein's unknown structure, based on its sequence, from fragments of related homologs. Even in cases of when the target protein has a known structure, a variation of comparative modeling may be necessary to generate alternative conformations to capture the inherent dynamics of the protein. Protein structure prediction is especially handy with membrane proteins that are more difficult to characterize than their soluble counterpart. The addition of modeling enables structural insights for entire protein families when only a few members have experimentally determined structures. A common workflow in these studies is to generate an ensemble of comparative models to represent a target receptor, and then docking a small molecule to the ensemble to identify possible binding modes.

4.2.2 Comparative modeling and docking with Rosetta

The Rosetta software suite, with its protein modeling and design foundation, is able to perform both comparative modeling and ligand docking. With the RosettaScripts ap-

plication system, the two protocols can be linked seamlessly. A full case example using T4-lysozyme is explored by Combs et. al[117]. In one benchmark Rosetta docking to comparative models, Kaufmann and Meiler observed a native-like binding pose among the top ten scoring for 21 out of 30 test cases. Furthermore, docking results were significantly better in cases when utilizing protein templates containing a ligand of similar chemotype compared to templates with dissimilar ligands or in the apo state[36]. In the case of G-protein coupled receptors, integral membrane proteins of great pharmaceutical interest, Nguyen et. al. found that RosettaLigand sampled near-native poses in docking into comparative models of 14 GPCRs, but it was challenging to select correct binding modes by score alone. The use of high sequence identity templates, knowledge-based binding pocket filters, and experimental contact points all served to improve accuracy[59].

4.2.3 Multi-template comparative modeling in Rosetta

Comparative modeling in Rosetta traditionally used a "copy and refine" approach. A given target sequence is matched with known structures from the Protein Data Bank (PDB), the matching segments are copied over, and short sequence-based fragment pieces are used to fill in the unmatched regions. Performance relied on having homologous proteins with at least 30% sequence identity. Proteins that maintained its tertiary fold and had well ordered structure were the easiest to model[1]. Proteins from less well characterized families were more challenging to predict. Furthermore, docking to comparative models becomes more difficult if the template protein structures were determined with no ligand or a chemically distinct ligand bound. This can be an issue even when the homolog template has high overall sequence identity with the target as the binding interface may not be conserved. One way to better comparative modeling in these cases is with RosettaCM, an improved sampling algorithm and scoring function that utilizes multiple templates with high local sequence identity.

RosettaCM[116] is a comparative modeling algorithm that assembles structures from

multiple aligned fragments. This allows for more accurate construction of a target protein even when there are no close homologs. In the CASP10 study, RosettaCM improved side-chain and backbone conformations when sequence identity was over 15%. A dedicated scoring function for an initial low-resolution fragment recombination stage favors compact structures with buried hydrophobic residues and paired beta strands. This initial stage generated structures of correct topology overall but are often distorted where fragments are joined. The second stage uses local fragment replacements to minimize individual areas before a third stage adds the full side chain representation of the protein. One particular area of improvement was in the modeling of loops connecting helices. This is particularly important to ligand docking as there is often an induced fit effect when a flexible loop is present near a ligand binding site.

4.2.4 STAT proteins

STAT, or signal transducer and activator of transcription, proteins are transcription factors that are activated by membrane associated receptors. All seven STAT proteins have been shown to play a role in disease with STAT3 and STAT5 being popular targets for small molecule inhibitors of human tumor growth[123, 124]. Berg et. al. previously discovered a potent STAT5 inhibitor with a naturally occurring chromone scaffold[125]. The question of whether a STAT5 inhibitor can be altered for activity against other STAT members was explored and tested using a fluorescence polarization assay[126].

4.2.4.1 Structure of erasin and STAT proteins

A derivative of the originally reported STAT5 inhibitor was found to have significant activity against STAT3 with lesser inhibition against STAT1. The compound, dubbed erasin, and its activity is shown in Figure 4.1. A major structural question is understanding why this compound exhibited different activities against the various STAT subtypes and if the compound can be rationally designed for desired selectivity.
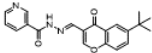
Structure	STAT1 app. IC_{50} [μM] or inhibition [%]	STAT3 app. IC_{50} [μM] or inhibition [%]	STAT5b app. IC_{50} [μM] or inhibition [%]
	23.0 ± 6.2	9.7 ± 1.8	$32 \pm 2\%$ inhibition at 80 μM

Figure 4.1: Activity of erasin in fluorescence polization assay against SH2 domains of STAT1, STAT3, and STAT5b.

Three crystal structures of relevance were identified in the PDB. a structure of unphosphorylated STAT1 complexed with a phosphopeptide, structure of the STAT3 homodimer bound to its DNA recognition site, and a structure of STAT5A that was a close homolog of the STAT5B structure. These structure was truncated to consider only the SH2 domain containing the ligand binding site, which is based on the peptide binding position observed in the STAT3 structure.

4.2.5 G-protein coupled receptors

GPCRs are transmembrane signal transduction proteins that comprise the largest human superfamily of receptors. They are characterized by seven transmembrane domains, and generally interact with either the cAMP or phosphatidylinositol signal pathways. A recent analysis shows 475 drugs, over a third of alll drugs approved by the FDA, target GPCRs. Existing drugs interact with 108 unique GPCRs with over 300 potential drugs in clinical trials that target either established or novel GPCR targets[127]. There are over 350 potential druggable GPCRs, i.e. GPCRs where one expects to be able to design a small molecule binder and the binding of the small molecule might alter the function with a therapeutic benefit. The discovery of new, selective GPCR ligands is also beneficial to studying GPCR function in biological systems[128]. This makes GPCRs one of the most critical targets for drug discovery[129, 130].

Only a fraction of GPCR-small molecule crystal structures have been solved across all four major classes (A,B,C,F). There remains over 100 orphan GPCRs, receptors for which

the endogenous small molecule is not yet known[129]. The GPCR Dock assessments reported near experimental accuracy for docking rigid orthosteric small molecules into close homologs, but docking flexible small molecules into binding pockets of distant homologs with flexible loops remains challenging. As existing structures cover only a fraction of druggable GPCRs, docking into comparative models will often be a necessary part of the modeling pipeline[131, 132]. The existing structures and associated structure activity relationship data provide excellent templates in driving modeling efforts[133].

4.2.5.1 Metabotropic gluatamate receptors

Metabotropic glutamate receptors (mGluRs) are a family of eight class-C mammalian GPCRs critical in excitatory activity in the CNS. The wide distribution of glutamatergic synapses suggests regulation of particular subtypes can be useful in treating a wide range of CNS diseases[134] including Fragile X syndrome[135], schizophrenia[136] and Parkinsons[137]. The eight mGluRs are further divided into groups I, II, and III based on their localization and function in the synapse.

Metabotropic glutamate receptor 3 is a group II member of the mGlu GPCR receptor family. Upon agonist binding, mGlu3 exerts cellular effects via inhibition of adenylyl cyclase. Activation of mGluR3 in turn modulates glutamate N-methyl-D-aspartate receptors, whose inhibition is known to induce schizophrenic behavior in otherwise healthy individuals. An mGluR3 agonist, LY354740, was shown to alleviate schizophrenia symptoms in an animal model[138]. Furthermore, genomic studies have identified single nucleotide polymorphisms in the mGluR3 gene associated with decreased pre-frontal cortex function, cognition, memory, as well as an increased risk of schizophrenia. In addition, the localization of group II mGlus in the forebrain and the differential distribution in presynaptic, postsynaptic, and glial compartments make them promising therapeutic targets if selective ligands can be developed.

4.2.5.2 Allosteric modulation of group II mGluRs

Previous development of allosteric modulators for group II receptors has been hindered by lack of selectivity between mGlu2 and mGlu3[139]. For instance, a previously disclosed NAM, LY2389575C[140], was shown to have a mere four-fold selectivity preference for mGlu3 over mGlu2[141] in a GPCR-potassium channel coupling thallium influx assay[142] . This intra-group preference is critical for therapeutic and probe development as the two subtypes have distinct mechanisms in neurodegeneration/neuroprotection associated with schizophrenia[143]. However, recent progress within the Vanderbilt Center for Neuroscience Drug Discovery has generated a novel mGlu3 selective NAM from a related mGlu5 PAM. The CNS-penetrant NAM, VU0463597, exhibits potent yet selective binding to mGlu3 (>15x over mGluR2) and has an SAR profile suitable for molecular docking experiments[141]. A library optimization strategy to improve upon the selectivity and DMPK profile of VU0463597 has yielded 410 total compounds, 137 of which are active in a calcium mobilization assay[119].

Figure 4.2 shows VU0463597 and examples of currently known structure-activity relationships. Two regions of particular interest are the MeO entity (green) implicated in selectivity over mGlu5, and the amide entity (teal) implicated in selectivity over mGlu2. Despite these advances, the underlying biological mechanism of mGlu3 allosteric modulation and selectivity over mGlu2/5 remains mysterious.

4.2.5.3 Crystal structures for mGlu receptors

There are two crystal structures for the transmembrane domain of mGluRs, mGlu1 and mGlu5, both in complex with a negative allosteric modulator[144, 145]. The mGlu1 structure, bound to the FITM ligand, was observed in a closed conformation. A N-terminus fusion with a thermostablized apocytochrome was performed for crystallization[145]. The mGlu5 structure was crystallized in complex with the mavoglurant negative allosteric mod-
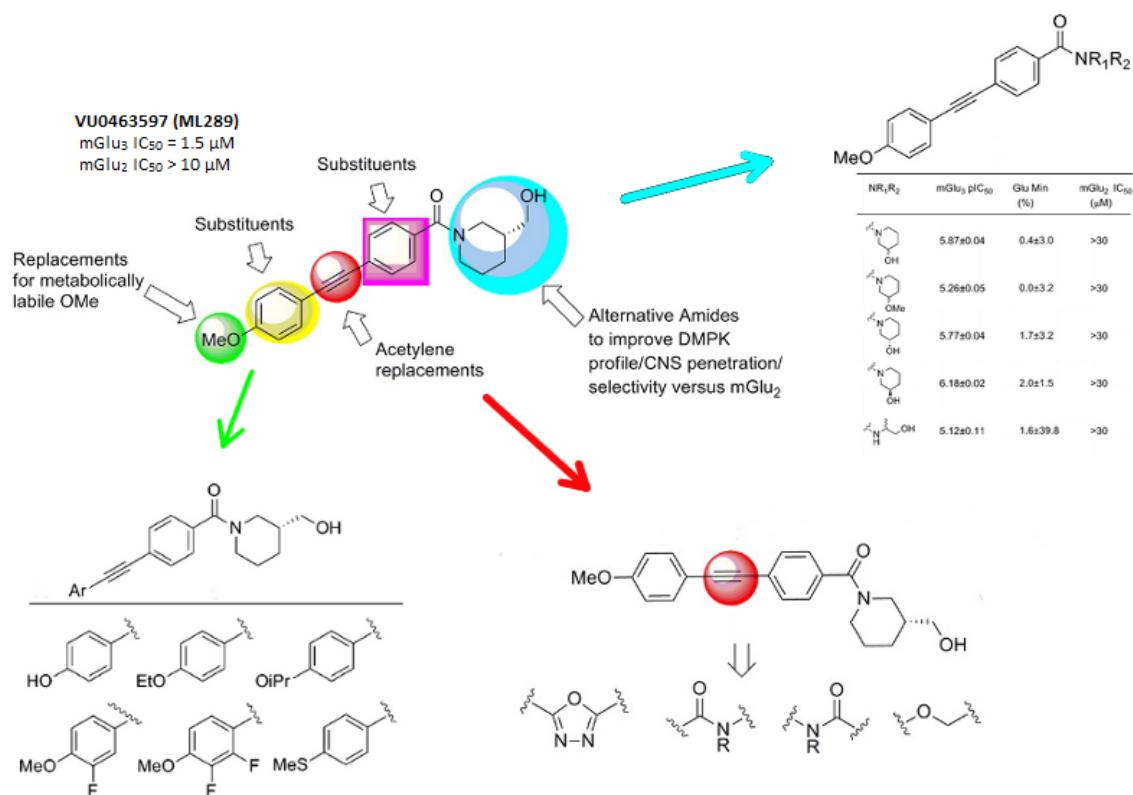
Figure 4.2: mGlu3 selective NAM VU0463597 is shown with critical chemical entities and examples of previously tested structural derivatives. Graphic reprinted from Wenthur et. al. 2013

ulator. A T4 lysozyme construct was inserted in the intercellular loop 2 to aid crystallization. Relative to the mGlu1 ligand binding site, the mavoglurant site was much deeper in the central helical core[144].

4.2.5.4 Protease activated receptors

Protease-activated receptors, or PAR, are a family of four (PAR 1-4) of GPCRs highly expressed in platelet cells. They are activated by thrombin signaling, a critical component of blood coagulation and thrombosis. Thrombin is known to cleave the N-terminus to reveal ligand tether region that binds to the second extracellular loop to initiate signaling. Several PAR4 antagonists are known to bind and block this binding site as a potential

antithrombotic[146]. However limited physiological understanding of PAR4 function have made drug discovery efforts challenging[147].

The structure of PAR1 in complex with the antagonist vorapaxar is available. Vorapaxar binding is highly specific but virtually irreversible, leading to its clinical limitations. The binding is very close to the extracellular surface but solvent exposure is limited due to the closure of the extracellular loop[148]. Recently, the structure of PAR2 in complex with two distinct antagonists have also been published. One of the antagonists exhibits slow binding kinetics and competed against the orthsteric tethered ligand, making it ideal for potential pharmaceutical development. The other antagonist bound in an allosteric pocket outside of the transmembrane domain, exerting its inhibitory influence by preventing the structural rearrangement necessary for PAR2 activation[149].

4.3   Experimental Methods

4.3.1   Modeling of target proteins

Rosetta comparative modeling was performed in each case using either a single template or multiple templates when available. Templates were restricted to proteins of the same family as at least one close homolog was available in each case. The template structures were available in the holo state with a ligand in the same putative binding pocket targeted by subsequent docking runs. For the GPCRs, a membrane protein specific scoring framework[150] was used in generating models.

4.3.1.1   STAT

The existing structures of STAT1 and STAT3 were energy minimized using the Rosetta FastRelax to generate an ensemble of models. STAT5b was constructed from the highly similar STAT5a homolog using single template "copy and refine" Rosetta comparative modeling.

4.3.1.2 mGlu3

An mGlu3 comparative model was onstructed using the RosettaCM protocol[116]. A sequence analysis of the 7TM domain shows mGlu3 is 45% identical to both mGlu1 and mGlu5. The mGlu1 and mGlu5 crystal structures exhibit 72% identity and 83% similarity over the 7TM domain to each other. The critical transmembrane helices and allosteric binding site are fairly well-resolved in the crystal structures. Gap regions in the crystal structure template are restricted to the intracellular and extracellular loops, which are relatively far away and unlikely to impact docking calculations. The same is true of the crystallization constructs observed in the mGlu1 and mGlu5 templates. Missing residues in the alignment can be filled in via cyclic coordinate descent[151]. In order to reduce bias, multiple top-scoring comparative models will be used for docking and the resulting scores will be averaged to select the best protein-ligand complex models.

4.3.1.3 PAR4

The existing structure of PAR1 in complex with vorapaxar is used to generate comparative models of PAR4 with single template modeling. This work is completed primarily by Alyssa Lokits but is unpublished. The top comparative models were used for docking. An effort to improve these models using RosettaCM with the PAR1 and PAR2 structures as multiple templates is ongoing.

4.3.2 Preparing and docking ligands

Ligand conformations are generated using the Molecular Operating Environment. RosettaLigand and RosettaLigandEnsemble docking was used with a low resolution shape complementarity stage and a high resolution knowledge based energy function to generate 1000 models. Models are selected based on "interface energy", an approximation for ligand binding energy calculated as the difference between the Rosetta score of the bound complex and

the separated protein/ligand molecules.

4.3.2.1   STAT

Erasin docking to STAT3 was performed allowing a full ligand reorientation. Prior
to interface energy evaluation, generated models were clustered using bcl::cluster with an
RMSD cutoff of 3 . The largest clusters were analyzed qualitatively and quantitatively to
determine possible binding interactions. To understand erasin's selectivity profile, STAT3
>STAT1 >STAT5b, models for erasin binding to STAT1 and STAT5b were also generated.
The position of erasin was aligned with pair fitting to the hypothesized binding mode from
STAT3 and a limited refinement with ligand and side-chain flexibility was performed.

4.3.2.2   mGlu3

Ligands are filtered to include only those with a defined biological effect (positive or
negative modulation) and IC50 values within detection limit. Compounds may be further
focused by evaluating their Tanimoto similarities to each other, and clustering molecules
that share the largest common scaffold. This will increase the likelihood that the selected
ligands adopt a common pose upon binding. A correlation weight of 500 was used with
RosettaLigandEnsemble to promote the generation of binding modes that match structure
activity relationships for the ligands. It is hypothesized that the mGlu3 ligands will bind in
a parallel position at the center of the transmembrane helices, similar to the pose adopted
by FITM and mavoglurant in the mGlu1 and mGlu5 structures. Prior to interface score av-
eraging, visual inspection was used to eliminate models where the ligand ensemble adopted
a perpendicular binding mode that protruded into the surrounding membrane. These likely
arose because the low resolution stage does not consider membrane nature of the protein
surroundings.

4.3.2.3   PAR4

A series of three BMS derivatives were docked with RosettaLigandEnsemble with a correlation weight of 500. It should be noted that experimental correlation is challenging in cases with few ligands since rank correlation takes on very discrete values. The binding mode of vorapaxar in PAR1 is used to make the initial placement of the BMS derivatives.

4.4   Results and Discussion

4.4.1   STAT

The fluorescence polarization assay to generate binding data is illustrated in Figure 4.3AB. The putative binding mode for erasin to STAT3 with key residues is shown in Figure 4.3C. Comparing the predicted STAT3 binding mode and STAT3 bound to a phosphotyrosine-containing peptide motif in Figure 4.3D shows significant conformation changes for STAT3 binding to erasin.

The sequence and structural ensemble used for docking are shown in Figure 4.4AB respectively. The STAT1 structure contains a highly flexible loop region away from the ligand binding site. The STAT5b structure was much more compact compared to STAT1 and STAT3. The binding modes of the three STAT proteins with erasin are shown in Figure 4.4C. Significant clashes exist between erasin and the STAT5b structure, and to a lesser extent, with the STAT1 structure as shown in Figure 4.4D. These clashes, shown in red and orange, are due primarily to intermolecular repulsive interactions scored on a Lennard-Jones potential and unfavorable protein dihedral angles scored by a knowledge-based potential. The acyl hydrazone part of erasin is located in a narrow channel. This is in accordance with SAR studies that have shown substitutions in this region has significant impact on binding.
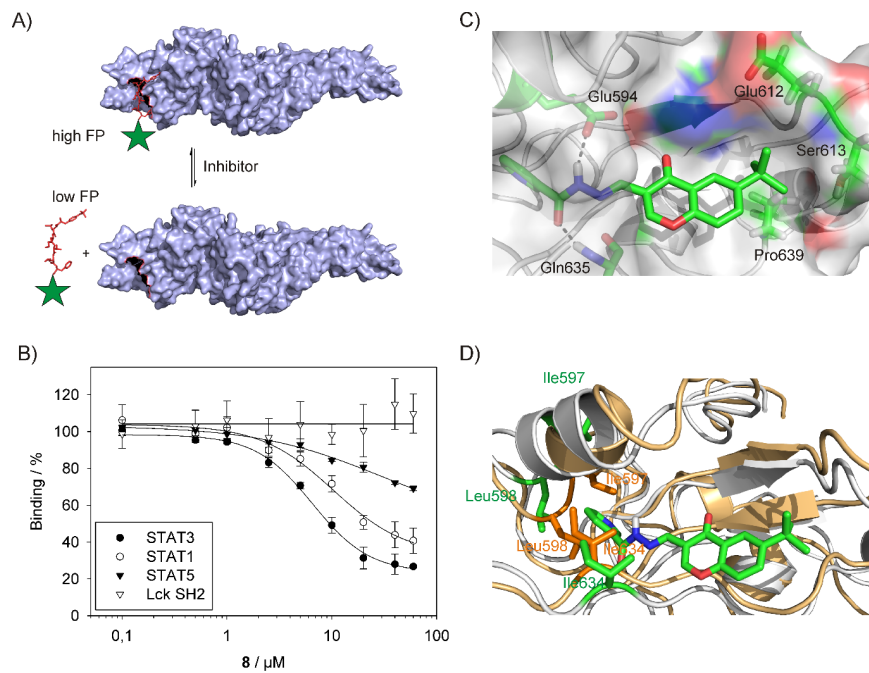
Figure 4.3: A: Principle of fluorescence polarization (FP)-based competitive binding assays. The fluorophore attached to the STAT SH2 domain binding peptide is indicated by a green star. B: Activity of erasin in fluorescence polarization assays against STAT1, STAT3, STAT5b and Lck. C: Docking pose of erasin. D) Overlay of the docking pose of erasin (STAT3 backbone shown in light grey, highlighted amino acid side chains shown in green) and the crystal structure of phosphorylated STAT3 (protein backbone shown in ochre, highlighted amino acid side-chains shown in orange, PDB 1BG1).

A) **STAT1**

```
WNDGCIMGFISKERERALLKDQQPGTFLL
RFSESSREGAITFTWVERSQNGGEPDFHA
VEPYTKKELSAVTFPDIIRNYKVMAAENI
PENPLKYLYPNIDKDHAFGKYYSR
```

STAT3

```
WNEGYIMGFISKERERAILSTKPPGTFLL
RFSESSKEGGVTFTWVEKDISGSTQIQSV
EPYTKQQLNNMSFAEIIMGYKIMDATNIL
VSPLVYLYPDIPKEEAFGKYCR
```

STAT5

```
WNDGAILGFVNKQQAHDLLINKPDGTFLL
RFSDSEIGGITIAWKFDSQERMFWNLMPF
TTRDFSIRSLADRLGDLNYLIYVFPDRPK
DEVYSKYYTPVP
```

B)

C)

D)

Figure 4.4: A) Amino acid sequence of the SH2 domains of STAT1, STAT3, and STAT5b. B) Structural models of the SH2 of STAT1, STAT3, and STAT5b in the absence of a ligand show regions of protein flexibility. C) Representative model of a cluster 3 binding pose of erasin bound to the STAT3 SH2 domain, and similar poses for STAT1 and STAT5b obtained by docking starting from the pose of erasin as observed in STAT3. D) Surface representation for models in panel C showing per-residue Rosetta energy units using rainbow color codes (red-orange-yellow-green-blue), with the least favorable interactions shown in red and the most favorable interactions shown in blue. Clashes (indicated in orange and yellow) do not occur in the binding pocket of erasin in STAT3, but to a small extent, in the equivalent region of STAT1, and to a large extent, in STAT5b.

4.4.2 mGlu3

Seven models were then selected out of the top twenty comparative models to account for conformational diversity. The selected models have up to 2.25 A of RMSD from the best scoring model in all TM residues and up to 1.5 A of RMSD in key residues. Key residues are defined as residues with at least one atom within 7 A of the allosteric binding site as suggested by the mGlu1 crystal structure.

RosettaLigandEnsemble was used to dock eight congeneric NAMs into the top comparative models for mGlu3. The congeneric NAMs were discovered by collaborators as derivatives of the mGlu3-selective lead VU0463597 shown in Figure 4.2. All modifications were made only to the amide scaffold in order to maintain selectivity for mGlu3/mGlu3. The selected compounds exhibited a ten-fold affinity range based on IC50 in a calcium flux assay. The best scoring model exhibited a rank correlation of 0.6 with experimental binding data and is shown in Figure 4.5.

Clustering analysis shows that the top scoring models generally adopted a binding mode wherein the MeO moiety in the compound faced towards the extracellular surface. The exact placement of the MeO group varied but interactions are frequently observed with the residues colored in yellow/teal. This is significant as the sequence alignment in 4.6 shows that the yellow residues are different in mGlu2/3 when compared with mGlu5. Previously published SAR studies on the ligand scaffold have indicated the MeO is necessary to eradicate mGlu5 binding. This model suggests that the R68Q change in mGlu5 may disrupt a key electrostatic interaction between the arginine and MeO, thus eliminating binding.

Furthermore, SARs indicate that substitutions at the amide moiety are critical in determining selectivity between mGlu2/mGlu3. However, it has been difficult to completely eradicate binding at one or the other. The orange residues shown interact with this critical amide end and are conserved between mGlu2 and mGlu3. The degree of conservation in the putative binding pocket explains why it remains challenging to select between the two group II mGlus. The model also identifies an important residue (D/N 168) that may be
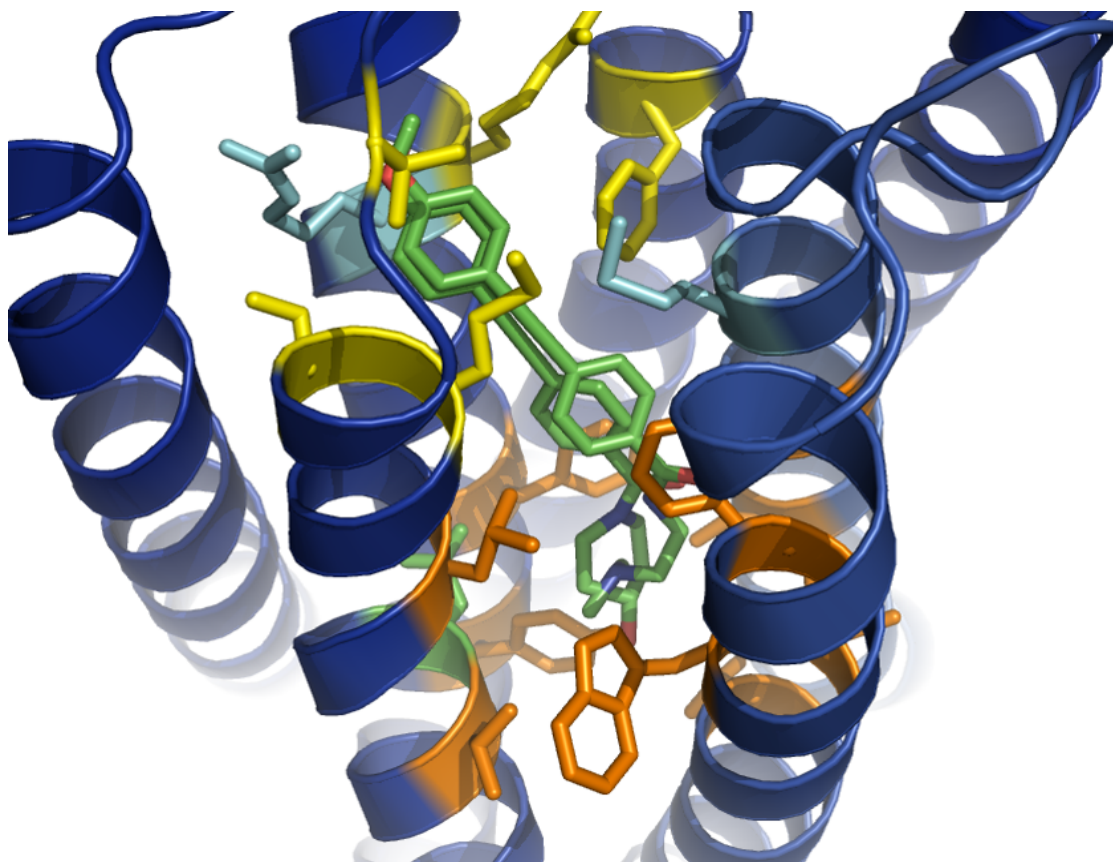
Figure 4.5: Top scoring model from RosettaLigandEnsemble docking of eight NAMs into seven mGlu3 comparative models. The view is from the extracellular side and key residues are color-coded according to the alignment in Figure 4.6.

specifically targeted in future ligand optimizations.
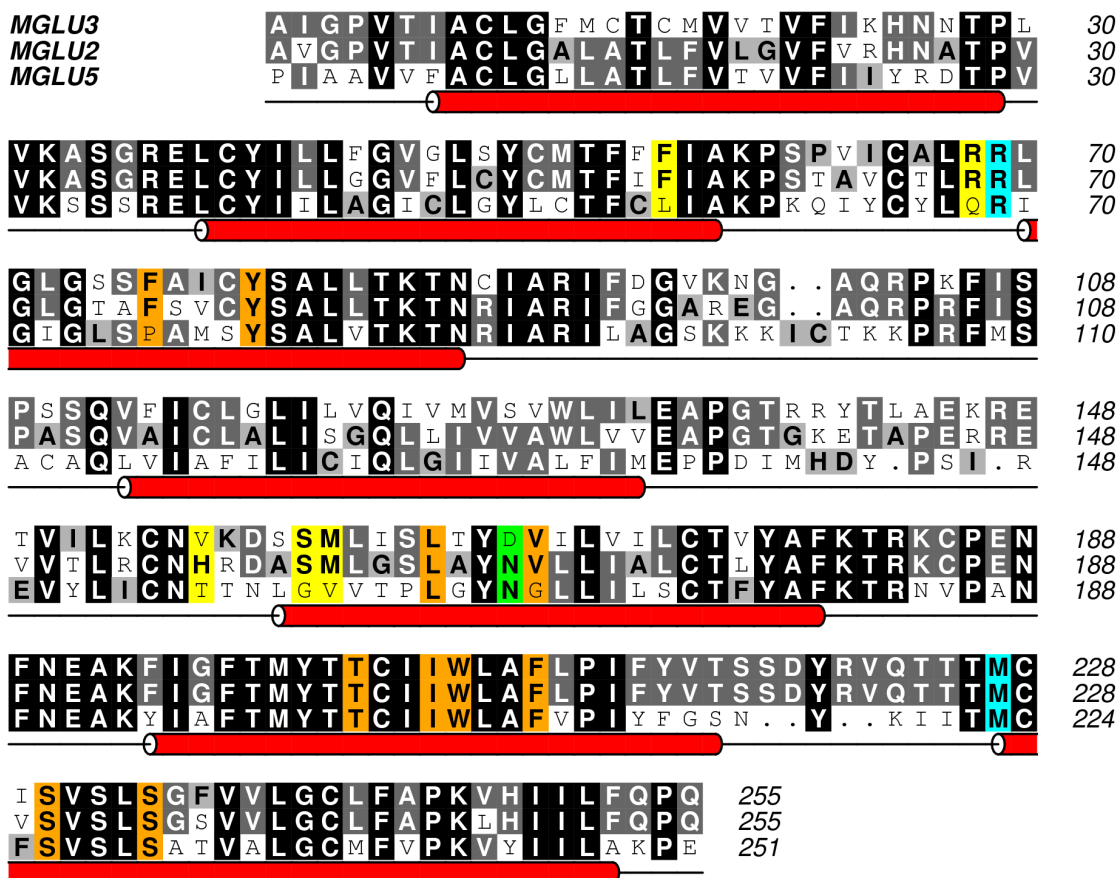
Figure 4.6: Sequence alignment of mGlu2, mGlu3, and mGlu5 crystal structures with the TM domain indicated by red helices. Key residues in the interface are colored as follows: residues different between mGlu3 and mGlu5 (yellow), different between mGlu3 and mGlu2 (green), same between mGlu3 and mGlu2 (orange), same among all three mGlus (teal).

4.4.3 PAR4

Figure 4.7 shows the docking of three derivatives with different sized alkyloxy substitutes into the comparative model of PAR4. The top scoring ensemble docking results show the alkyloxy group pointing into the center of the helical core. The docked ensemble matches the experimental data that the largest group, the benzyloxy, had the strongest biological inhibition. This suggests some sort of anchoring mechanism for the ligand. The conservation coloring shows a limited number of possible mutations for explaining the

PAR1 vs PAR4 selectivity of these compounds.
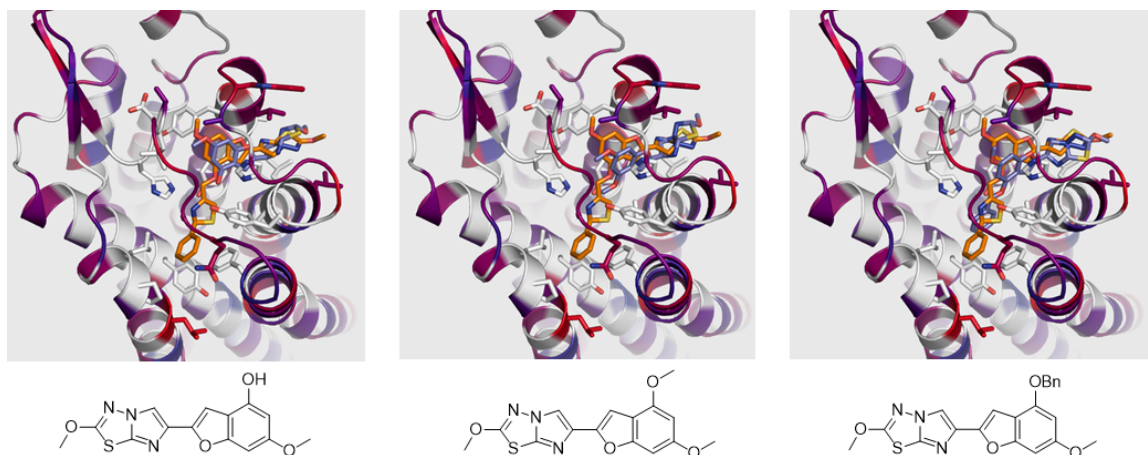
Figure 4.7: Docked model of hydroxyl (left), methoxy(center), and benzyloxy (right) derivatives are shown in blue. Vorapaxar position in PAR1 is shown in orange. Protein residues are colored by sequence similarity to PAR1 with white being identical and mutations are colored from least conserved (red) to most conserved (purple).

4.5 Conclusion

4.5.1 STAT

Rosetta docking of erasin to STAT1, STAT3, and STAT5b generated a possible binding mode that explains the protein and ligand structure activity relationships. Further studies are needed to examine the protein structural cause of the binding selectivity. One hypothesis is the lack of a key interaction on STAT5b that causes a loop to close over the narrow channel observed in STAT3. Experimental mutagenesis studies would be needed to confirm this hypothesis. A work currently under review examines the flexibility exhibited by these loops, and their impact on the binding site.

4.5.2 mGlu3

Rosetta was used to generate mGlu3 models from mGlu1 and mGlu5 templates. RosettaLigandEnsemble was successfully used to dock a congeneric series of NAMs into these

comparative models to find a binding mode that matches the structure activity relationships. These models are used to create hypotheses that could be tested using a double mutant cycle. In particular, mutation of critical residues in mGlu3 to their mGlu1 or mGlu5 counterpart should eliminate binding for a mGlu3 selective ligand.

4.5.3 PAR4

RosettaLigandEnsemble was used to dock BMS antagonists into a comparative model of PAR4. Docking generated vorapaxar like binding modes with a portion of the molecule extending deep into the helical core. Possible mutagenesis sites were selected for testing but have not yet been experimentally validated. The project is ongoing and has a focus on updating PAR models and docking a new chemical class of inhibitors.

Chapter 5

Conclusion

5.1 Summary

Chapter 1 discussed the general background of protein-ligand docking. The central theme of this dissertation is the improvement of protein-ligand docking tools by the addition of structural ensembles and structure-activity relationships. The chapter discusses existing methods for incorporating experimental data and critically examines the assumptions behind molecular similarity. The included analysis showed that it is generally sound to presume that congeneric small molecules bound in highly similar fashions to a given protein target.

Chapter 2 presented a new docking method, RosettaLigandEnsemble, that allows for the use of a multiple ligand overlay to simultaneously find binding modes for related small molecules. It is designed to work in conjunction with SAR studies that provide many modifications to a central scaffold. There was a significant improvement in sampling efficiency over the existing RosettaLigand algorithm, particularly for ligand series with significant modifications to certain members. However, generating binding scores in correlation with the SAR remained challenging due to the stochastic nature of the docking algorithm and the steepness of the energy well around the native ligand binding mode. This chapter is paired with Appendix B which contains a tutorial for running RosettaLigandEnsemble.

Chapter 3 covered the upgrades to an automated ligand docking server. The goal of this project was to improve the accessibility of Rosetta docking methods. In particular, it would be beneficial for those without any computational biology training to be able to use Rosetta for docking and modeling. However, there are still additional features to be added to ROSIE ligand docking such as a simplification of input file formats and a clearer

interpretation of output results.

Chapter 4 applied RosettaLigand and RosettaLigandEnsemble to collaborative small molecule discovery projects. Docking was used in conjunction with Rosetta comparative modeling, a critical protocol for use cases without a readily available crystallographic structure of the protein target. It also touched on the pharmacological importance of G-protein coupled receptors and the computational techniques necessary to approach GPCR ligand docking. The STAT, mGlu3, and PAR applications all involved exploring the question of how a small molecule binds and the structural factors that explain its selectivity. Although only one of the projects was successfully completed and published at the present time, the applications show the potential of RosettaLigandEnsemble docking as part of the drug discovery pipeline.

Method development would not be possible without the proper benchmarking datasets. Appendix A covers the datasets used to optimize the protocols within this thesis. These datasets are made available as a lab resource for future research.

A proof of concept study considering structural ensembles from the protein side is demonstrated in Appendix . This was a significant method development towards protein-ligand docking incorporating structure-activity data from both ligand modifications and protein mutations. The new method, ProtLigEnsemble, will require an extensive benchmark, beyond the scope of this dissertation, to fully understand its strengths and weaknesses. The protocol for running ProtLigEnsemble is included in this appendix.

5.2   Key findings

The introduction of biochemical data in the modeling process is a strong promoter of protein-small molecule modeling success and accuracy. Spectroscopic methods such as NMR and mass spectrometry have been adapted to interrogate receptor-small molecule interactions. One can expect improvements in both sampling and scoring when incorporating experimental contacts or structure activity relationships. Structure activity relationships de-

rive from energetic changes upon modifications to either the receptor or the small molecule. Docking improvements are particularly significant when experimental restraints limit both the translational and the rotational modes of the small molecule. Such restraints can be derived from interface-based experiments that identifies specific protein-ligand interactions, or by relying on binding modes of similar small molecules. These methods would rely on the assumption that similarities in structure translates to similarities in binding. A pairwise comparison across the PDBBind database shows that this is largely true among congeneric small molecules. Exceptions exist wherein typically highly symmetric molecules bind in inverted orientations. These situations may be addressed by further development of pharmacophore, or molecular property, based alignment methods.

RosettaLigandEnsemble docking significantly improved the sampling efficiency in a benchmark set of congeneric ligand structures. However, additional pre and post processing is necessary to use the algorithm. Correlation weighting is capable of producing sets of structures that match the SAR data, however current scoring functions are not accurate enough for this to be a significant benefit. Applying the existing scoring function to the native-like binding mode is unlikely to rank the ligands in the correct order.

Applications of RosettaLigand and RosettaligandEnsemble to mGlu3, PAR, and STAT proteins generate reasonable models of protein-ligand interactions. The predicted binding mode identifies several key contacts to explain observations of ligand SARs and protein mutant selectivity. Further experimentation is necessary to validate the proposed binding mode. One issue is the resource commitment necessary to make mutagenesis studies to test the binding mode. This can be challenging if the goals of the computational approach do not match those of the experimental approach. In particular, an exhaustive synthesis approach with readily available ligand fragments may be preferred over limited synthesis based on probabilistic computational predictions.

5.3 Future Outlook

Although the research advanced the central theme of accurate protein-ligand modeling in the context of structure-based drug discovery, there are a number of considerations that need to be addressed further. Two in particular are how to address the consistently more challenging task of ranking small molecules, and to bridge the gap between computation and traditional "wet-lab" chemistry.

5.3.1 The challenge of small molecule scoring

Although the work in thesis focused primarily on docking to recover binding modes, the major challenge remains the ability to accurately score, or rank, small molecule ligands. Docking assessments have consistently found ranking to be the harder of the two tasks. An improvement to scoring is likely to carry with it an improvement to native pose recovery. As the D3R benchmark illustrated with its MAP4K4 docking benchmark, many submissions had a native like binding prediction among its top 5 scoring models but only half identified it as the most accurate model. Furthermore, the best correlation with experimental affinity obtained for ranking was 0.48[25].

One potential way to improve ligand scoring is through the use of system specific scoring functions. This is particularly useful for protein systems such as kinases where there is an ever growing wealth of data. Ross et. al. demonstrated that an universal scoring function, trained across a diverse data set, is less accurate than a target-specific scoring function when applied to a single system. A score function that captures experimental affinities across the entire dataset has varying accuracy on single protein datasets[152]. Thus, protein target dependence remains a significant challenge in choosing diverse benchmark targets for scoring function development, and in selecting a score function for use in a particular application system. Current approaches to overcome this bias are to test a multitude of scoring protocols before selecting one, or to utilize rescoring algorithms

such as NNScore that can be suited to a specific receptor for screening applications[153]. Trained scoring functions such as the random forest based SFCscore or the support vector machine based SVRR may be fitted to a given target by careful choice of training data and/or descriptors[154, 155]. One major advantage of tuned scoring functions is the ability to reproduce experimental activity data. AutoShim, for example, uses provided IC50 activity data and partial least squares regression to parameterize the scoring function. This empirical correction improved experimental SAR correlation from an all-purpose scoring function best of 0.32 to 0.5 across a GSK docking set[156]. However, care must be taken to avoid overfitting as having the same protein families in training and validation can produce unrealistically high prediction accuracy[157].

5.3.2 Applications to virtual screening

Ligand scoring and ranking is also fundamental to the task of screening. This applies both in a broad high throughput virtual screening situation where an algorithm needs to score interactions across a variety of molecular scaffold, and in a local optimization scenario where an algorithm need to be able to rank small modifications to a functional group. Although this work focused on binding mode recovery, an natural extension would be to the task of virtual screening. DeLuca et. al. [18] created tools for using the low resolution shape complementarity function for doing rapid virtual screening. Rosetta has also been used in a raycasting docking approach called DARC for targeting protein-protein interaction sites[158]. Ensemble docking can be used to perform virtual screening against multiple receptor targets. Although a number of methods currently use this approach for screening against multiple conformations of a flexible protein, a new application would be to screen against multiple protein mutants.

Computational approaches for multi-target screening is useful for addressing at least two basic research problems. First, certain mutations will invalidate the activity of an existent drug. This is the case for many target proteins with a high mutation rate such as

HIV proteins. The problem of binding erasing mutations is addressed from the protein side with multi-target antibodies. A similar approach could be done from the small molecule approach. Secondly, screening against multiple targets can be used to find multi-target drugs as a replacement for combination therapy. In these cases, targets may not necessary be the same protein binding target but rather a different target altogether[159]. It has been suggested though that the diversity of protein binding pockets may be represented by a thousand shapes[160], making it entirely possible to find multi-target drugs by screening against multiple protein targets aligned by pocket similarities.

One additional improvment in this area is the advent of big data. Machine learning is already a part of other aspects of the drug discovery process such as predicting organic synthesis reaction routes and outcomes[161]. As previously discussed, machine learning is also contributing to ligand docking in the development of system specific scoring functions to bolster ranking capabilities. Big data driven approachs can be applied to target discovery, particularly in the context of a limited number of binding pockets. This can be used for drug repurposing existing compounds, saving time and effort on the front of safety and synthesis testing [162]. This is particularly true given the ever growing size of compound bioactivity databases. One challenge however is that the ever growing database of activity data do not all reflect the same level of confidence in the bioassay[163]. There is certainly still a need for structural based machine learning approaches to drug discovery.

5.3.3 The accessibility of computational predictions

Docking as a predictive tool is only useful if it can be coupled with traditional wet-lab experimental verification. A binding mode is merely a hypothesis until it can be matched to experimental data. Receptor or ligand side data can bolster support for a model but coupled interface data such as double mutant cycles can provide stronger validation. One gap to bridge is the mismatch in goals between wet-lab benchwork and dry-lab in silico work. Computational approaches often come with the idea that no question is too basic.

The ability of docking to successfully predict a protein-ligand complex structure is not a given. However, experimental groups often have much further goals for a computational collaboration. This came into play during the mGlu3 project. My work generally focused on trying to determine the binding of presently available compounds while our collaborators were interested in the development of DREDDs, or designer receptors exclusively activated by designer drugs. Docking to GPCRs, in particular a yet to be crystallized class C GPCR, was still a challenging problem and interface design was many steps away.

Part of this gap is due to a fundamental misunderstanding regarding the capabilities and the timescale of computation. Computational methods are often described in collaborations without the subtle nuances needed to understand their limitations. This is partially due to poor communication between the computational and benchwork sides. One way to foster better communication is increasing the availability of computational methods for non-computational scientists. Many collaborations feature computational and wet-lab biologist working independently but coming together to share data. It could be beneficial to develop easy to use tools for those without computational backgrounds. In particular, most researchers are familiar with "point-and-click" user interfaces in other scientific software. It should be an standard in the computational field to offer graphical ways to operate the methods. Furthermore, visualization of structural data is not always straight forward and often require its own dedicated software tools or plugins. An easier way to share graphical data and its interpretations could go a long way towards understanding what a computational prediction entails and means.

Appendix A

Description of Datasets

This appendix describes the various datasets constructed for the research presented in this thesis. The datasets can be accessed in the Meiler Lab's internal storage.

A.1   RosettaLigandEnsemble Congeneric Ligands Validation Set

A dataset of 109 protein-ligand complexes across twenty systems A.1 are curated from the combination of the Community Structure-Activity Resource [164], BindingDB Protein-Ligand Validation Sets[90], PDBBind[165], D3R docking resource, and individual crystallographic studies[166, 167, 98].

Each dataset consisted of at least four chemically related ligands with experimental data and X-ray crystallography determined structures against a common protein target. A single receptor structure was selected from each dataset as the primary docking target on the basis of crystallographic resolution, density in the ligand binding pocket, and experimental affinity/activity. In order to test the potential of an ensemble docking approach, the dataset favors cases wherein congeneric ligands bind in a similar fashion and an improvement using RLE docking is expected. Figure A.2 shows the distribution of congeneric ligand RMSDs and common scaffold sizes seen in the dataset.

The selected protein receptor structure is energy minimized using the Rosetta FastRelax protocol with a knowledge based all-atom energy function[170]. The details of the Rosetta energy function has been covered extensively by Alford et. al[19]. This minimization is performed in the apo state to remove bias of side-chain positioning for the co-crystallized ligand. All other molecules in the series are cross-docked to the energy minimized target using either traditional RosettaLigand docking or simultaneous RLE docking. Ligand

conformations are generated using the in-house BioChemicalLibrary fragment-based con-former sampling methodology[95]. The co-crystallized ligand is excluded from docking with RosettaLigand and RLE to avoid any bias, leaving a total of 89 test cases across all systems.

Protein System	Name	Count	Target	Source
Androgen receptor ligand binding domain	AR	5	3B66	BindingDB
A. vaga glutamate receptor 1 ligand binding domain	AVGLU	5	4IO6	PDBBind
C. fimi xylanase Cex	BETAX	5	1FH9	PDBBind
T. gondii calcium-dependent protein kinase 1	CALM	5	3V51	PDBBind
Bovine spleen cathepsin B	CATB	6	2DCB	BindingDB
Cyclin dependent kinase 2	CDK2	6	2VTL	BindingDB
P53 core domain mutant Y220C	CTAP	6	4AGQ	PDBBind
Human coagulation factor Xa	FXA	4	1NFW	BindingDB
M. tuberculosis malate synthase	GLCB	5	3S9I	Krieger 2012
Hepatitus C virus NS5B polymerase	HCV	6	3CO9	BindingDB
Heat shock protein 90	HSP90	6	4YKT	Gathiaka 2016
UDP-3-O-N-acetylglucosamine deacetylase	LPXC	5	4FW7	CSAR
H. pylori MTAN nucleosidase	MTAN	6	4FFS	PDBBind
Biphenyl amide p38 kinase	P38	4	3D83	BindingDB
Phenylethanolamine N-Methyltransferase	PNMT	5	2G72	PDBBind
Protein tyrosine phosphatase 1B	TPPHO	6	2QBQ	CSAR
Human retinoic acid nuclear receptor	RET	4	1FCY	PDBBind
Spleen tyrosine kinase	SYK	8	4YJT	CSAR
Bacterial thermolysin	THERM	7	4N4E	Krimmer 2014
Human thrombin	THROM	5	2ZDA	Baum 2010

Figure A.1: Protein-Ligand systems used to benchmark RosettaLigandEnsemble Sources, ligand counts, receptor PDBs, and abbreviations of the twenty benchmark systems. Each protein-ligand set contains crystallographic structures of one receptor in complex with each ligand. The systems are derived from BindingDB (https://www.bindingdb.org/bind/index.jsp)[168], PDBBind (www.pdbbind.org)[169], Community Structure-Activity Resource (CSAR, http://www.csardock.org/)[164], and individual studies as listed.
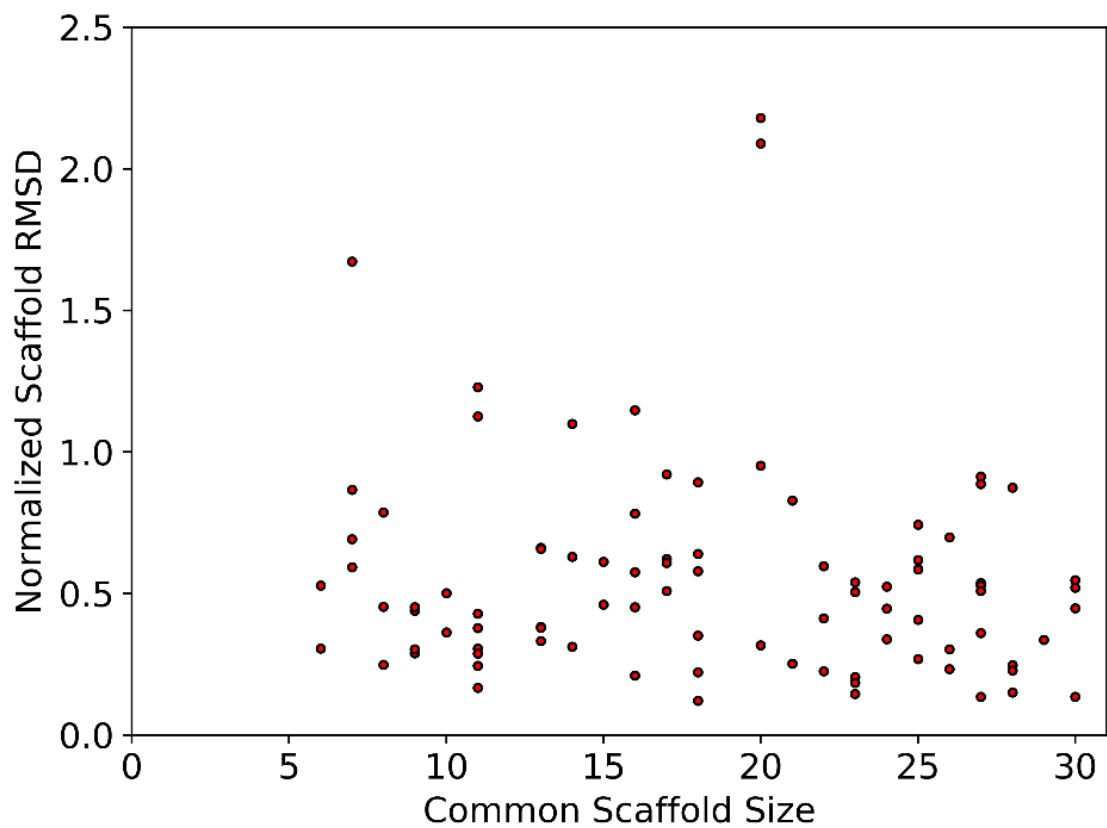
Figure A.2: Number of non-hydrogen atoms in the common scaffold versus the RMSD (normalized to 15 heavy atoms) of the common scaffold. Each molecule is compared to the ligand co-crystallized with the receptor structure for the system it belongs to.

A.2 PDBBind Core Set

The PDBBind Core Set is a subset of the PDB Refined Set that contains 65 targets with 3 ligands each. Each triplet contains a low affinity binder, a medium affinity binder, and a high affinity binder. A subset of 65 protein-ligand complexes is constructed using only the high affinity binding molecules[27]. For benchmark purposes, the protein receptors are relaxed using Rosetta FastRelax[170] in the apo state to reduce bias due to crystallographic side-chain positioning.

A.3 PDBBind Refined Set

The PDBBind refined set contains 3446 structures controlled for protein structure quality, accurate binding data, small molecule properties, and non-surface interactions. Each protein-small molecule co-crystal structure is paired with experimentally measured Kd, Ki, or IC50[27, 169]. In order to analyze similarities in congeneric ligand binding, systems with only one crystal structure were fitlered out. This produced 2443 structures across 441 targets. All complexes within a system were then aligned based on binding pocket residues within 15 of the small molecule. The in-house BioChemicalLibrary (BCL) software suite, available at http://www.meilerlab.org/bclcommons, is used to calculate small molecule properties and make all possible intrasystem pairwise comparisons. The pairwise normalized RMSD (nRMSD) is calculated based on the heavy atom RMSD of the largest common connected substructure and normalized to ten heavy atoms using a small molecule analog of RMSD-100[171]. The pairwise Tanimoto similarity coefficient is computed as the number of atoms in the largest common connected substructure divided by the total number of unique atoms.

All 34461 pairwise comparisons were then filtered to eliminate identical small molecule pairs (Tanimoto = 1), trivial common scaffolds (scaffold heavy atoms <5), and different binding pockets (small molecule center distance >3.0). A large number of the remaining

comparisons were from the HIV-1-Protease system, which dominates the PDBBind refined set. In order to avoid biasing the results towards any particular system, only a randomly chosen small subset of HIV-1-Protease comparisons are included. Small molecule symmetry was factored in when calculating nRMSDs for pairs where symmetrical molecules are flipped. The final subset included a total of 7298 comparisons across 366 targets. The median Tanimoto similarity across the dataset was 0.333 and the median nRMSD of the common scaffold was 1.071 .

A.4   GPCR Modeling and Docking Set

A GPCR modeling and docking set was constructed in order to test the effectiveness of docking ligands into comparative models of GPCRs. A set of sixteen GPCR structures were collected from the PDB. Each GPCR would be modeled with RosettaCM[116] using the remaining fifteen as templates. Comparative modeling would focus primarily on the transmembrane regions with loop building used to reconstruct intracellular and extracellular loops. Figure A.3 shows the transmembrane domain sequence identity (lower triangle) and C-alpha RMSDs (upper triangle). This dataset need to be updated as newer GPCR protein-ligand complex structures become available.

	2RH1	3ODU	3PBL	3UON	3VW7	4EIY	4IAR	4IB4	4JKV	4K5Y	4N6H	4OO9	4OR2	4PHU	4PXZ	4S0V
2RH1		1.58	1.22	1.23	2.06	1.58	1.08	1.28	2.09	1.99	1.34	2.31	2.07	1.81	1.88	1.14
3ODU	18.0		1.56	1.68	1.84	1.89	1.71	1.70	1.87	2.23	1.41	2.39	2.46	1.95	1.88	1.59
3PBL	29.0	20.0		1.32	2.00	1.39	1.31	1.33	2.00	1.99	1.39	2.08	2.00	1.91	2.11	1.19
3UON	26.0	18.0	29.0		2.22	1.71	1.42	1.65	2.12	2.24	1.50	2.36	2.43	2.17	2.44	1.30
3VW7	17.0	22.0	18.0	19.0		2.13	1.75	1.87	2.42	2.36	1.90	2.72	2.22	1.52	1.58	1.95
4EIY	27.0	15.0	26.0	23.0	15.0		1.52	1.55	2.22	1.99	1.79	2.28	2.30	2.13	2.30	1.57
4IAR	32.0	19.0	34.0	31.0	16.0	26.0		1.32	2.14	2.22	1.43	2.45	2.30	1.75	1.93	1.18
4IB4	28.0	18.0	30.0	26.0	16.0	24.0	30.0		2.12	1.81	1.72	2.33	2.41	1.78	2.23	1.45
4JKV	8.0	8.0	7.0	8.0	9.0	7.0	6.0	6.0		1.96	2.10	2.14	2.01	2.19	2.30	2.20
4K5Y	9.0	10.0	8.0	10.0	10.0	11.0	10.0	9.0	12.0		2.16	2.16	2.15	2.24	2.28	1.90
4N6H	22.0	22.0	22.0	23.0	22.0	21.0	23.0	22.0	7.0	12.0		2.31	2.25	1.73	1.84	1.31
4OO9	8.0	8.0	6.0	7.0	7.0	9.0	10.0	7.0	6.0	7.0	8.0		0.67	2.33	2.57	2.35
4OR2	8.0	8.0	7.0	8.0	6.0	10.0	7.0	9.0	7.0	7.0	9.0	70.0		2.14	2.19	2.12
4PHU	14.0	18.0	13.0	13.0	20.0	13.0	11.0	16.0	10.0	9.0	16.0	11.0	10.0		2.04	1.84
4PXZ	15.0	21.0	17.0	17.0	19.0	13.0	16.0	16.0	8.0	10.0	21.0	6.0	7.0	15.0		2.06
4S0V	23.0	23.0	24.0	21.0	19.0	24.0	22.0	20.0	6.0	14.0	24.0	9.0	8.0	13.0	16.0	

Figure A.3: Transmembrane domain sequence similarity among the sixteen benchmark GPCRs is shown in the lower triangle. Aligned C-alpha RMSDs over the same transmembrane region is shown in the upper triangle. Red indicates a higher similarity while blue indicates greater differences.

Appendix B

Protocol Capture for RosettaLigandEnsemble

B.1 Obtaining files

Rosetta can be obtained through www.rosettacommons.org All files associated with this protocol capture is provided in the demos/protocol_capture/rosettaligand_ensemble/ directory of the Rosetta distribution. This protocol has been tested to work with Rosetta version d978e6f, released August 22nd, 2017. Examples commands for this protocol are numbered in the commands file of the protocol capture folder and referenced as (1), (2), (3)etc.

B.2 Starting files

The raw starting files are a single target protein receptor structure in PDB format, and a series of ligands in SDF format. The protein structure can be prepared from an experimentally determined structure or from homology modeling. The receptor structure used in this example is the p53 core domain bound to a stabilizing small molecule (PDB: 4AGQ). This file can be found in /inputs/ as protein.pdb. The ligand series should share a core scaffold by which the ligands can be aligned. This example contains five congeneric ligands, but any number between three and eight is a reasonable use case.

B.3 Ligand preparation

PyMol pair fitting is an easy way to manually align ligands by minimizing the distance between core scaffold atoms. Automated ligand alignment tools may also be used but generally do not perform as well compared to manual inspection. Examples of aligned ligands can be found in the /prep/aligned_ligands/ directory.

Each ligand must have its own conformational library generated prior to using Roset-taLigandEnsemble. Conformer generation was done using the BioChemicalLibrary. The BCL (http://www.meilerlab.org/servers/bcl-academic-license) is a suite of software tools readily available for academic users. Other software for conformer generation may also be used but the outputs need to be converted to SDF files.

To generate conformers using the default settings, use command (1). Conformer generation can also be customized to use the PDB or CSD libraries, a greater range of rotamers, or a structural comparison filter to remove similar conformers. For a full list of these options, see the help menu with command (2). Examples of generated conformers files are in /prep/conformers/

B.4   Param file preparation

Rosetta requires params file to properly handle small molecule ligands. Prior to this step, join the aligned ligand structure with the corresponding conformers into a single SDF file such that the aligned structure is first in the file. This will insure that the inputs will maintain the core scaffold alignment when generating the conformers. Examples of these joined files are in /prep/make_params/

Since PDB files use three digit residue codes and single digit chain designations, it is helpful to assign a code for each ligand file. The example uses the ligands.list file to label each ligand as residues 00B through 00F and corresponding chains B through F. This file also contains pK values for each ligand binding to the target receptor.

To make params files for each ligand, use command (3). The molfile_to_params python script is included in the /main/source/scripts/python/public/ directory of the Rosetta distribution. Running the script without any input prints out the help menu.

For each ligand, command (3) generates a PDB file containing the single aligned ligand structure, a PDB file containing the remaining ligand conformers, and a params file containing connectivity and charge information for the ligand. Examples of these files can

be found in the /prep/rosetta_inputs/ directory using the previously discussed letter designations. If you wish to incorporate SAR during docking, then use a text editor to add NUMERIC_PROPERTY AFFINITY YOURVALUE to the end of the params file, where YOURVALUE represents an SAR measurement of the users choice. Rank correlation is used in SAR mode and hence units only need to be self-consistent. RosettaLigandEnsemble assumes that smaller affinities are more favorable just as smaller Rosetta scores are considered more favorable. The behavior can be adjusted for cases where larger affinity values is more favorable by either taking the negative of the provided affinities in the params file, or by making the correlation weight negative.

B.5   Input file organization

For RLE runs, it is convenient to prepare a single PDB file containing the aligned ligands by concatenating the individual ligand PDB files. The conformer and params do not need to be joined. This is done as ligands.pdb in the /inputs/ folder. Youll also find the previously prepared protein receptor PDB in the same directory.

In addition to structural files, a RosettaScripts XML file and a Rosetta options file. The XML file describes the custom protocol to be used by Rosetta. Details of how to setup an XML file and the meaning of the individual tags can be found by searching the documentation website https://www.rosettacommons.org/docs/latest/. The example dock.xml provided uses the settings from the benchmark. Actual application use may require the user to alter these values according to biological context. The defined scoring function is based on the existing RosettaLigand scoring function, but may be substituted in the XML script. The provided options file defines Rosetta input and output directories along with a number of sampling parameters. A full options list is available on the documentation website. The ligand_ensemble option is necessary to use RLE; a weight of 0 can be used to run RLE without taking SAR data into consideration.

Run command (4) to perform a single simulation and generate a set of RLE models.

Each simulation will produce X models, where X is the number of input ligands. These example output models are in the /outputs/ directory along with a score.sc scorefile.

B.6   Output and analysis

Individual protein-ligand predicted structures are labeled by a chain and a number designation, B_1.pdb through F_1.pdb. Structures with the same numeric label are based on the same docking simulation and have a common binding pose. The protein interface contacting each ligand are optimized independently. The score.sc file contains all score terms for each simulation across a single row. Generally, individual ligand interface scores are used to rank models, with a negative score indicating a better model. These ligand interface scores are listed as interface_delta_*, where * is the single letter ligand chain ID. The values are appended at the end of each output PDB, and also in the scorefile for each protein-ligand pair. One suggestion is for the end user to examine the top ten percent of models for each pair.

Appendix C

Additional Developments of Ensemble Docking

C.1   Summary

This appendix describes ongoing research on the development and benchmark of ensemble docking approaches that merge structure-activity relationships derived from ligand modifications and protein mutations. A new algorithm, ProtLigEnsemble, was created to use such SAR data and a proof of concept study was completed. The basis of the algorithm and the proof of concept will be submitted as a Rxiv preprint for which I will be the sole first author. A full peer reviewed manuscript is expected once a more extensive benchmark has been completed. This appendix includes a protocol capture illustrating how the new Rosetta based ensemble features should be used.

C.2   Ensemble approaches from the protein perspective

RosettaLigandEnsemble was developed to dock a series of congeneric ligands to a single protein target with the option to include structure-activity relationship (SARs) for ligand modifications. As one might imagine, this can be extended to utilize the SARs associated with protein changes rather than ligand changes. As discussed in Chapter 1, ligand binding changes upon protein residue mutagenesis is often utilized as a way to localize the interaction site. We have extended ligand docking in Rosetta to process mutation data automatically and to use the SARs to guide modeling.

Existing methods make use of protein ensembles for working with challenging protein targets that are not well represented by a single static structure. This often occurs with protein targets with highly flexible binding sites and/or a significant induced fit effect. In one study, Ellingson et. al. used molecular dynamics snapshots to improve decoy dis-

crimination over docking against a single crystal structure[172]. This strategy can also be used to generate the holo protein conformation when starting with an apo structure [173]. Although Rosetta does not perform molecular dynamics, we have now enabled the ability to use protein ensembles with an "average-grid" scoring method in RosettaLigand and RosettaLigandEnsemble. This greatly increases the amount of protein flexibility Rosetta can process on top of the backbone and side-chain degrees of freedom typically allowed. One interesting feature is that Rosetta protein ensembles do not have to be comprised of the same protein variant. In other words, Rosetta protein ensembles can be used to dock or screen against multiple mutants simultaneously.

C.3   Combining protein and ligand SARs

A further combination for ensemble docking would be the use of protein-based and ligand-based SAR. This modality is geared towards cases where a ligand panel is tested against several protein mutants. Figure C.1 shows the distinction among the datasets for each of the ensemble docking approaches.

RosettaLigandEnsemble is discussed in Chapter 2 while protein ensemble docking and protein ligand "double" ensemble docking is new to this appendix. Protein ensemble docking can be performed by enabling additional options in the existing Rosetta ligand docking movers. Double ensemble docking represents a novel method, ProtLigEnsemble, which will be detailed further in this appendix. ProtLigEnsemble only requires a single input structure either from experimental determination or comparative modeling. The remaining structures will be generated by residue subsitution and energy minimization as the general overall fold should remain the same. The presence of misfolding mutations can be experimentally determined without the need for a fully solved structure. Missing binding data in the table would be ignored by the algorithm.

RosettaLigandEnsemble

	WT
Ligand 1	Structure Binding
Ligand 2	Binding
Ligand 3	Binding

Protein Ensemble Docking

	WT	Mutant 1	Mutant 2	Mutant 3
Ligand 1	Structure Binding	Binding	Binding	Binding

Protein Ligand "Double" Ensemble

	WT	Mutant 1	Mutant 2	Mutant 3
Ligand 1	Structure Binding	Binding	Binding	Binding
Ligand 2	Binding	Binding	Binding	Binding
Ligand 3	Binding	Binding	Binding	Binding

Figure C.1: Experimental datasets that different ensemble approaches are designed to work with. In each case, it is assumed that the ligands have a common central scaffold and that binding data is measured the same wway for all protein-ligand pairs. Only a single input structure is required and it does not have to be the wild-type protein.

C.4   Ensemble docking workflows

The new ensemble features can be utilized as part of three workflows summarized in Figure C.2. Workflow A represents RosettaLigand single protein-single ligand docking protocol described in DeLuca et. al.[18]. The modification now allows alternate receptor conformations to be passed into the low resolution docking phase with the ensemble_proteins XML tag. The alternate receptor conformations or mutations are used to calculate the scoring grid in the same manner as the original protocol. A single protein-ligand complex is produced using either the best scoring receptor model or the user's preferred receptor model.

Workflow B is the multiple ligand-single protein protocol described in Chapter 2. Similarly to workflow A, a new feature allows alternate receptor models to be passed into the low resolution docking phase. A single receptor model is passed on to the high resolution

docking phase where SAR guided ligand ensemble docking takes place. Users can specify whether they wish to use the best scoring model or to use the primary input model with the use_main_model XML tag. The final output contains one protein receptor in complex with each of the input ligands.

Workflow C is the new double ensemble method for multiple proteins-multiple ligands docking. The approach in docking is similar to that of RosettaLigandEnsemble in the low resolution phase. Alternate conformations may also be used in this stage. The additional protein mutants are generated in the high resolution ProtLigEnsemble stage. Similar to RosettaLigandEnsemble, an experimental dta rank correlation will be used to promote docking modes that generate the most favorable Rosetta scores for the strongest binders. Compounds without binding data to a given receptor will not be factored into the Spearman's coefficient. The final output contains a structure for every possible pair of input ligands and input proteins.

Additional changes in ProtLigEnsemble are the optimization order and the docking radius. The RosettaLigandEnsemble SAR guidance optimized one protein-ligand pair after another in iterative cycles. This has been adjusted to complete all optimization cycles for the protein-ligand pair with the highest experimental affinity first. This should generate the best scoring structures for the strongest binders, improving correlation with experimental data and reducing the stochastic nature of the output model scores. Furthermore, a docking radius has been introduced as a simple way of defining the binding pocket. Previous iterations required individual definitions for each ligand, leading to increasingly bulky XML scripts. The new version automatically defines the flexible portion of the protein structure based on a single setting and applies it to all protein-ligand pairs.
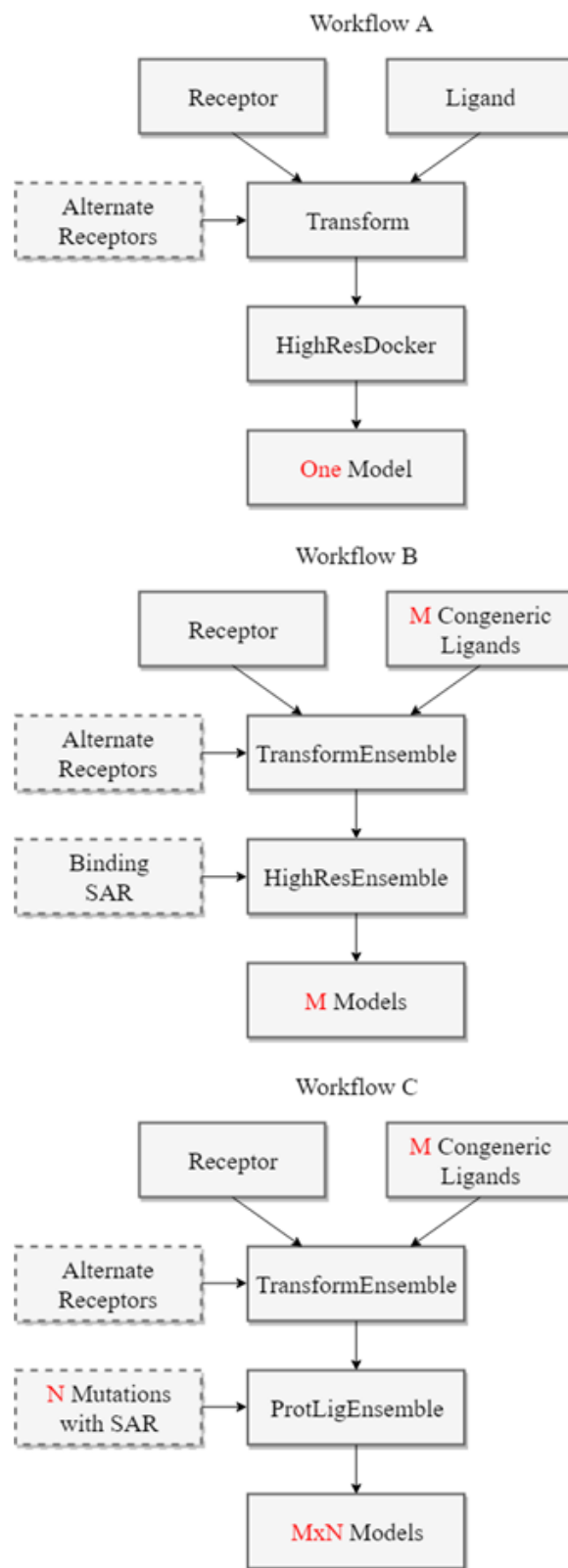
Figure C.2: Suggested workflows with optional inputs shown in broken line boxes. A: RosettaLigand workflow for using multiple receptor structures. B: RosettaLigandEnsemble workflow for using multiple receptor structures and SAR refinement. C: Protein ligand double ensemble workflow to incorporate both receptor- and ligand-based SAR.

C.5 ProtLigEnsemble proof of concept dataset

Two test cases are derived from individual studies on two GPCRs, the adenosine A2A receptor and the neuropeptide Y1 receptor. The A2A dataset focus on work by Zhukov et. al. [174] in which a panel of antagonists are tested against point mutants of a thermostablized receptor.
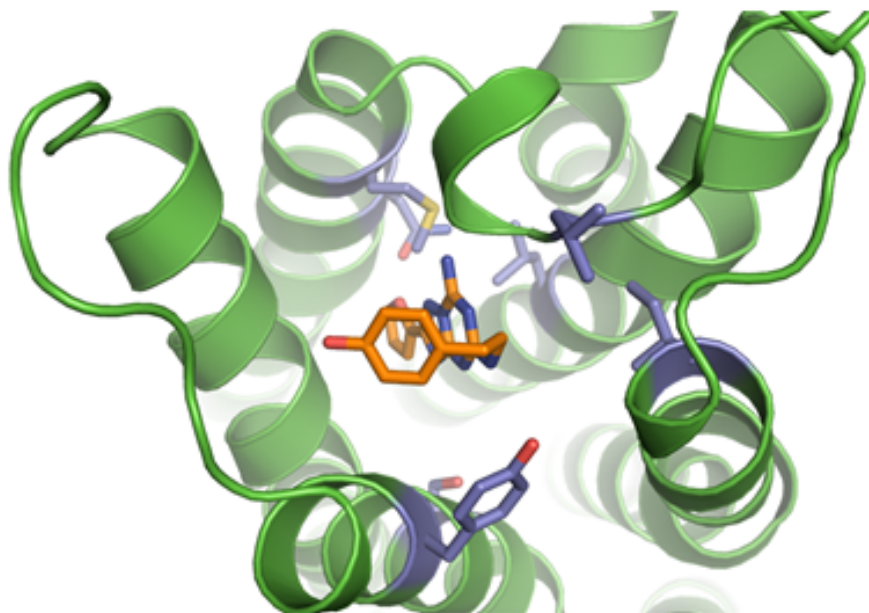
	ZM241385	SCH420814
Wild Type	9.3	9.5
I 66 A (2.64)	8.7	8.5
L 85 A (3.33)	7.5	7.2
L 167 A (ECL2)	9.3	8.5
M 177 A (5.38)	9.2	8.6
N 181 A (5.42)	8.4	8.3
S 277 A (7.42)	8.9	8.6

Figure C.3: Experimental dataset derived from Zhukov et. al. containing binding data (pKd) for two ligands against seven receptor variants. Crystal structure of wild type receptor in complex with ZM241385 (PDB: 3EML) is shown in green, the ligand in orange, and the mutation sites marked out in purple.

Figure C.3 shows the binding pocket of the A2A-ZM241385 complex with available mutational data denoted in purple. SCH420814 is a ligand with a significant addition to the extracellular binding end of the ligand. The crystal structure was published separately by Jaakola et. al. (PDB: 3EML)[175].

The Y1 dataset, shown in Figure C.4 is derived from Yang et. al., which contains a crystal structure of the Y1 receptor in complex with the antagonist UR-MK299 (PDB: 5ZBQ)[176]. The additional antagonists feature the addition of various linear amide groups to the central scaffold of UR-MK299.

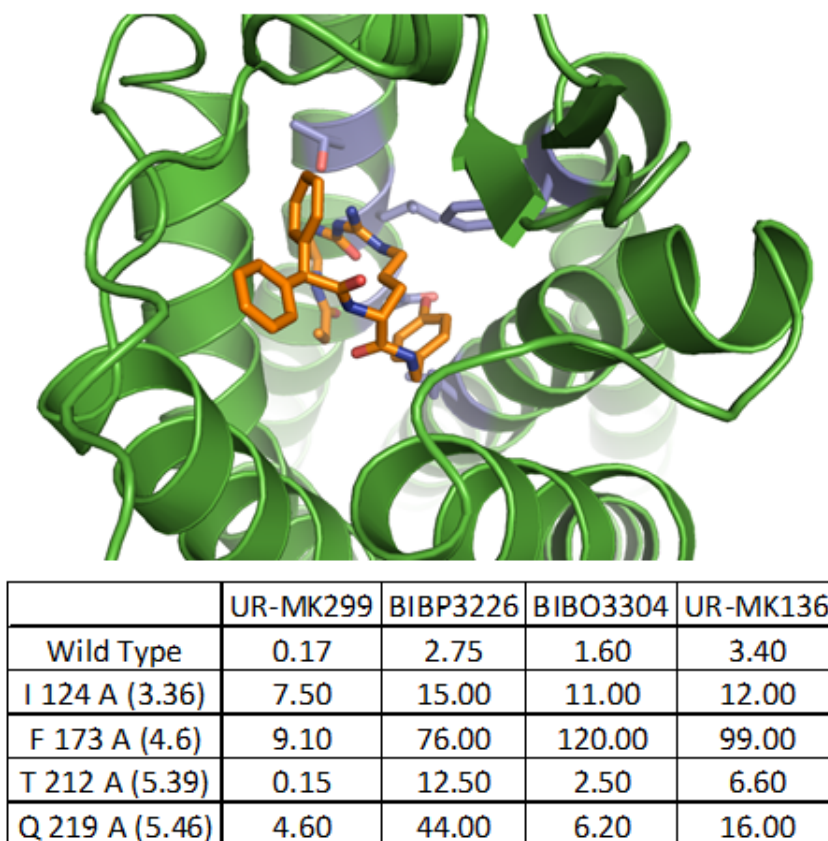|	UR-MK299	BIBP3226	BIBO3304	UR-MK136
Wild Type	0.17	2.75	1.60	3.40
I 124 A (3.36)	7.50	15.00	11.00	12.00
F 173 A (4.6)	9.10	76.00	120.00	99.00
T 212 A (5.39)	0.15	12.50	2.50	6.60
Q 219 A (5.46)	4.60	44.00	6.20	16.00

Figure C.4: Experimental dataset derived from Yang et. al. containing binding data (nM Kd) for four ligands against five receptor variants. Crystal structure of wild type receptor in complex with UR-MK299 (PDB: 5ZBQ) is shown in green, the ligand in orange, and the mutation sites marked out in purple.

C.6 ProtLigEnsemble docking results

Figure C.5 summarizes the docking results for the A2A test case. The docked models capture the general orientation of the binding mode observed in the crystal structure as indicated by the significant portion of docked models around 4 RMSD. However, the compact conformation of the ZM241385 ligand with the wildtype receptor is not well captured. This may be an issue with the pregenerated conformations rather than the docking method. One issue with the RMSD analysis is that only one protein-ligand pair can be compared for native-like binding modes. Alternative measures may be necessary to evaluate the efficacy of the method across the entire dataset. A significant enrichment of binding modes that correlate positively with the SAR is observed for the A2A test.
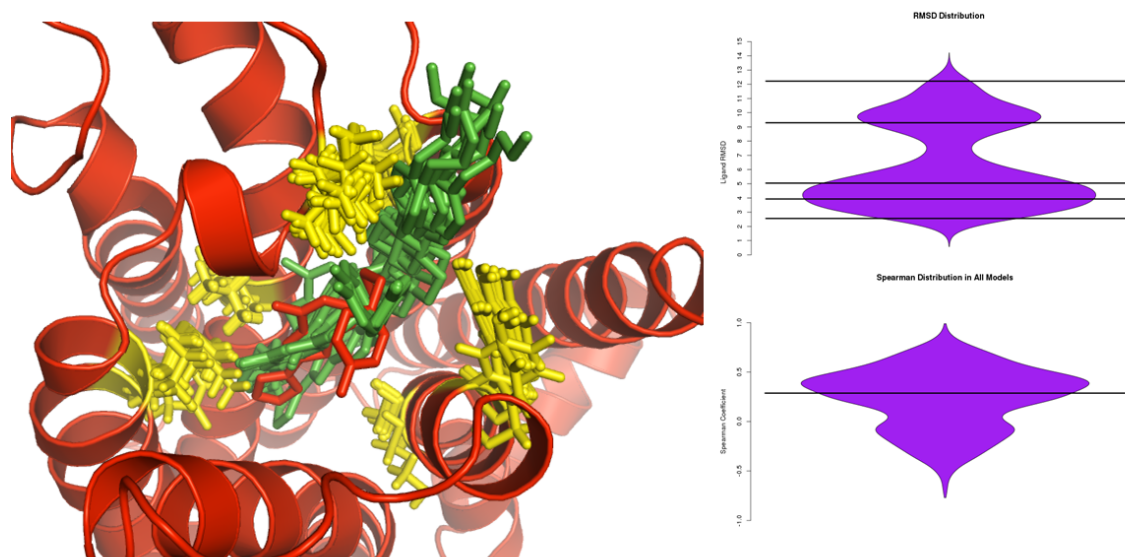
Figure C.5: Summary of docking results with top right graph showing the RMSD distribution of wildtype-ZM241385 models, and the bottom right graph showing the experimental rank correlation. A representative output is shown overlaid with the 3EML crystal structure. The crystal receptor structure and bound ligand is in red, docked ligands are in green, and mutation sites are shown in yellow.

Figure C.6 shows the docked models from a single simulation in grid format. One benefit of double ensemble docking over individual docking runs is the consistency of docking modes across the entire system. Rather than performing ad-hoc analysis to distinguish pu-

tative binding modes generated from individual runs, ProtLigEnsemble identifies binding modes consistent with the entire dataset. The interfaces consists of individually optimized ligand conformations and protein side-chain positioning. One limitation of this dataset is that the side-chain mutations consisted of alanine scanning, which in some cases simply represent a reduction in side-chain steric volume. A more extensive dataset including significant electrostatic based mutations could be useful in testing how well ProtLigEnsemble captures the SAR of substantial changes.
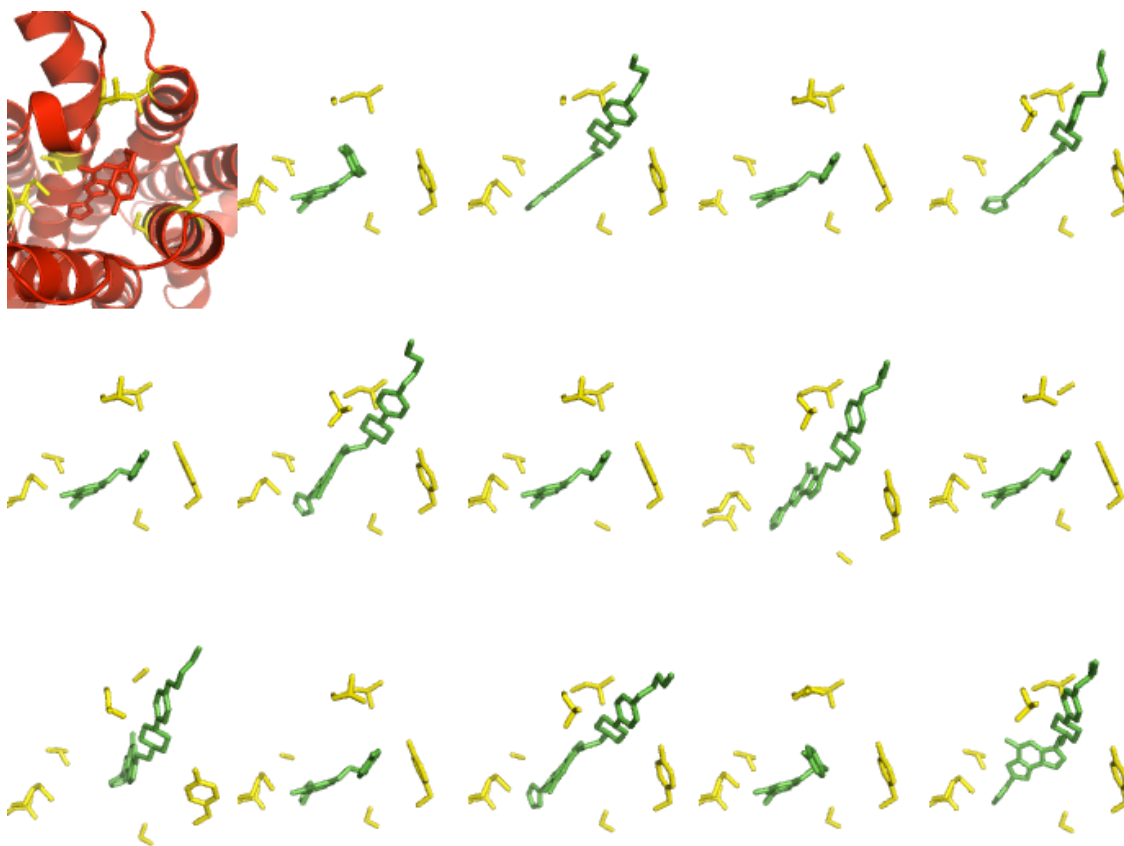
Figure C.6: Docking output from a single run of double ensemble docking. Top left corner shows the crystal structure of wild-type protein in complex with ZM241385. Remaining panels show each of 14 possible combination of 2 ligands with 7 protein variants. The crystal structure is in red, docked ligands are in green, and mutation sites are shown in yellow.

Figure C.7 shows the Y1 docking results summary. As indicated by the RMSD distribution, the native like binding mode was captured for a significant number of wildtype-

UR-MK299 models generated. One issue that continues to arise is the band of high RMSD models. Like RosettaLigandEnsemble, ProtLigEnsemble does not take into account SAR during the low resolution docking phase. Future development could focus on a way to improve model selection with low resolution scoring grids without adding substantial computational cost. The Spearman distribution in the Y1 test case is also less than ideal. This is likely due to the larger number of protein-ligand pairs being modeled. ProtLigEnsemble optimizes binding pairs in order of affinity and it is possible that the SAR correlation becomes too restrictive when docking the weaker binding pairs. This is analogous to the difficulties of negative design when using an algorithm designed to find more and more favorable scores.

Figure C.7: Summary of docking results with top right graph showing the RMSD distribution of wildtype-UR-MK299 models, and the bottom right graph showing the experimental rank correlation. A representative output is shown overlaid with the 5Z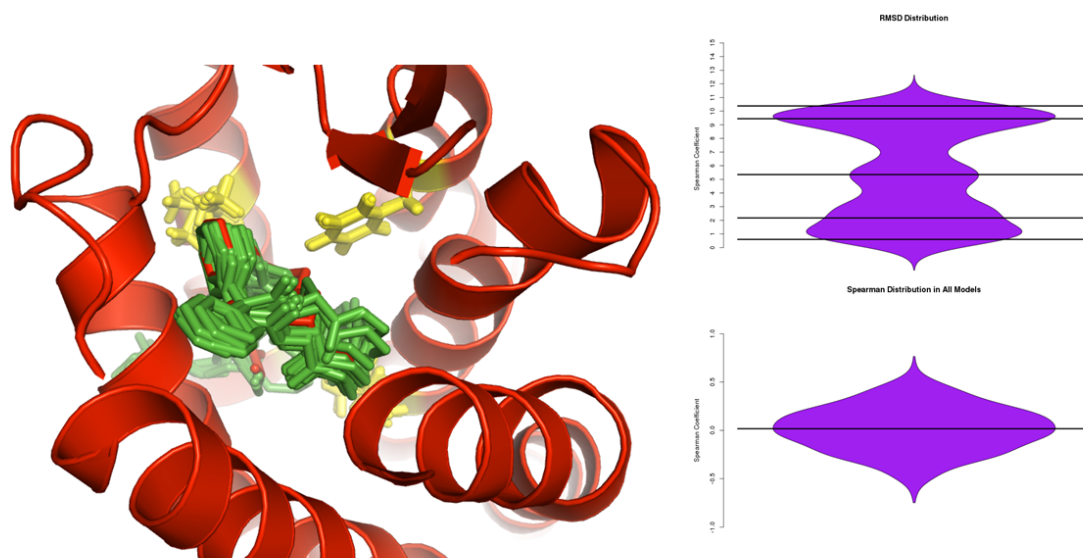BQ crystal structure. The crystal receptor structure and bound ligand is in red, docked ligands are in green, and mutation sites are shown in yellow.

C.7   Conclusion and future directions

Protein ensemble and double ensemble docking have been incorporated into the Rosetta ligand docking suite. The proof of concept shows that ProtLigEnsemble is capable of generate a consistent set of docked models with protein-ligand SAR guidance. Additional benchmarking is necessary to explore the full potential and limitations of the new algorithms. One suggestion for benchmarking the addition of protein ensembles to the RosettaLigand and RosettaLigandEnsemble workflows is via PSCDB. PSCDB is a database that tracks different types of receptor motion upon ligand binding[177]. This may be particularly helpful in assessing the use of ensemble methods for ligand binding to highly flexible loop regions.

To augment the double ensemble proof of study benchmark, the GPCR dataset described in Appendix A may be a good starting point. Many of these receptor studies include mutational data in their supplemental and there may be additional ligand compounds related to the co-crystallized ligand. It is unlikely a binding value for each protein mutant - ligand pair will be found, but ProtLigEnsemble can work around this limitation by ignoring missing SAR values. One other concern for dataset building is that different assays may be used for different protein-ligand combinations. In particular, some of the ligand data may be a measure of biological response, which can be due to allosteric effects as opposed to binding affinity changes. Careful curation will be necessary to generate a proper and diverse benchmark dataset.

C.8   ProtLigEnsemble protocol capture

The following section provides an example for using the double ensemble docking protocol. Rosetta can be obtained through www.rosettacommons.org

All files associated with this protocol capture is provided in the demos/protocol_capture/rosettaligand_P directory of the Rosetta distribution. This protocol has been tested to work with Rosetta

version 9820fea, released July 12, 2018. Examples commands for this protocol are numbered in the commands file of the protocol capture folder and referenced as (1), (2), (3)etc.

C.8.1   Starting files

The raw starting files are a single target protein receptor structure in PDB format, and a series of ligands in SDF format. The receptor structure used in this example is neuropeptide Y1 receptor bound to the ligand UR-MK299 (PDB: 5ZBQ). This file can be found in /inputs/ as protein.pdb. The four congeneric ligands in /prep/aligned_ligands/ directory have been aligned by their core scaffold. The reasonable number of ligands depends on the number of protein variants considered as each run generates all possible pairs. We generated up to 30 models per docking run without issue though this number may change depending on your computational setup.

C.8.2   Ligand preparation

Ligand preparation can be performed in the same fashion as the procedure in Appendix B. In particular, commands 1, 2, and 3 cover the process of generating conformations with the BCL conformer generator and creating Rosetta ligand param files. Example ligand conformers have been created for you in /prep/conformers and the necessary param generation files are provided in /prep/make_params/

The final Rosetta input ligand files are provided in /prep/rosetta_inputs/. The ligands have been designated with the letters B,C,D,E though you are free to use any chain designation as long as they are different from each other and the protein receptor. The correspondence between published ligand designations and the Rosetta lettering is provided in ligands.list. The params file process is the same as those for RosettaLigandEnsemble but you can skip adding SAR data to the param files as SAR data will be provided in a separate dedicated file.

C.8.3 Setting up the QSAR file

The QSAR file, an example of which is provided as inputs/qsar.txt, will provide protein-ligand pairs of interest to the ProtLigEnsemble mover. Each line is organized as a protein identifier, a ligand identifier, and an optional binding value. To provide a ligand binding value to a wild-type protein, enter:

WT B 0.17

where WT indicates wildtype, B is the single letter chain of the ligand, and 0.17 is the optional affinity value. Note that the affinity value can be any measure as long as they are self consistent. Rosetta assumes the lower values indicates a more favorable binding. This can be changed by setting a negative correlation weight in the -docking:ligand:ligand_ensemble option. To provide a binding value to a mutant protein, enter:

107 A B 7.5

where 107 A indicates a mutation at residue 107 to alanine, B is the single letter chain of the ligand, and 7.5 is the optional affinity value. This will cause Rosetta to generate a 107A mutant regardless of what the wildtype residue at position 107 is. Note that this numbering system must correspond to Rosetta pose numbering, where the first residue is numbered 1, the second is numbered 2...and so on. For the time being, ProtLigEnsemble is designed to work with single mutants. This QSAR file will be provided to Rosetta in the XML script as the qsar_file tag for the ProtLigEnsemble mover.

C.8.4 Input file organization

For ProtLigEnsemble runs, it is preferred to combine all aligned ligand PDBs into a single PDB file. This is provided as ligands.pdb in the /inputs/ folder. Note that the params and conformers files are not combined, just the single ligand PDB inputs.

In addition to structural files, a RosettaScripts XML file and a Rosetta options file. The XML file describes the custom protocol to be used by Rosetta. Details of how to setup

an XML file and the meaning of the individual tags can be found by searching the documentation website https://www.rosettacommons.org/docs/latest/. The example dock.xml provided uses the settings from the benchmark. Actual application use may require the user to alter these values according to biological context. The defined scoring function is based on the existing RosettaLigand scoring function, but may be substituted in the XML script. The provided options file defines Rosetta input and output directories along with a number of sampling parameters. A full options list is available on the documentation website. The ligand_ensemble option is necessary to use ProtLigEnsemble; a weight of 0 can be used to run ProtLigEnsemble without taking SAR data into consideration.

A few XML tags in the ProtLigEnsemble mover are newer features to this mover. The aforementioned qsar_file tag tells Rosetta where to find the QSAR file. The distance tag defines the radius, in angstroms, of the sphere around the ligand considered to be the binding pocket. All residues in this sphere are considered to be flexible. This tag replaces the LigandArea, InterfaceBuilder, and MoveMap tags RosettaLigand users may be familiar with. The ignore_correlation option tells ProtLigEnsemble to avoid calculating the rank correlation until there are at least 4 protein-ligand pairs in the dataset. This is because ProtLigEnsemble optimizes binding pairs in order of binding affinity. Considering the rank correlation with only a few protein-ligand pairs is not particularly useful. This option may be adjusted based on the number of ligands in your particular dataset.

Run command (4) to perform a single simulation and generate a set of ProtLigEnsemble models. Each simulation will produce X models, where X is the number of protein-ligand pairs listed in the SAR file. These example output models are in the /outputs/ directory along with a score.sc scorefile.

C.8.5   Output and analysis

Individual protein-ligand predicted structures are labeled by a protein-ligand pair designation. Wildtype proteins and ligand combinations will be tagged as WT_B_1.pdb through

WT_E_1.pdb where the 1 indicates the docking run it came from. Mutant receptor-ligand pairs will be tagged as 107_A_B_1.pdb through 107_A_E_1.pdb. The first two parts indicate the mutant residue number and the mutant residue identity respectively. These are followed by the ligand chain designation and the docking run number.

Structures with the same numeric label are based on the same docking simulation and have a common binding pose. The protein interface contacting each ligand are optimized independently. The score.sc file contains all score terms for each simulation across a single row. Generally, individual ligand interface scores are used to rank models, with a negative score indicating a better model. These ligand interface scores are listed as interface_delta_*, where * corresponds to the protein-ligand prefix tag seen in the PDB files. The values are appended at the end of each output PDB, and also in the scorefile for each protein-ligand pair. One suggestion is for the end user to examine the top ten percent of models for each pair.

BIBLIOGRAPHY

[1] Brian J. Bender, Alberto Cisneros, Amanda M. Duran, Jessica A. Finn, Darwin Fu, Alyssa D. Lokits, Benjamin K. Mueller, Amandeep Kaur Sangha, Marion F. Sauer, Alexander M. Sevy, Gregory Sliwoski, Jonathan H. Sheehan, Frank DiMaio, Jens Meiler, and Rocco Moretti. Protocols for Molecular Modeling with Rosetta3 and RosettaScripts. *Biochemistry*, 55(34):acs.biochem.6b00444, 2016.

[2] Darwin Y. Fu and Jens Meiler. Predictive Power of Different Types of Experimental Restraints in Small Molecule Docking: A Review. *Journal of Chemical Information and Modeling*, 58(2):225–233, feb 2018.

[3] M J Barnes. FUNCTION OF ASCORBIC ACID IN COLLAGEN METABOLISM. *Annals of the New York Academy of Sciences*, 258(1):264–277, sep 1975.

[4] E. M. Wise and James T Park. Penicillin: its basic site of action as an inhibitor of a peptide cross-linking reaction in cell wall mucopeptide synthesis. *Proceedings of the National Academy of Sciences*, 54(1):75–81, jul 1965.

[5] George Scatchard. the Attractions of Proteins for Small Molecules and Ions. *Annals of the New York Academy of Sciences*, 51(4):660–672, 1949.

[6] Subha Kalyaanamoorthy and Yi-Ping Phoebe Chen. Structure-based drug design to augment hit discovery. *Drug Discovery Today*, 16(17-18):831–839, 2011.

[7] C R Beddell, P J Goodford, F E Norrington, S Wilkinson, and R Wootton. Compounds Designed to Fit a Site of Known Structure in Human Hemoglobin. *British J. Pharm.*, 57(2):201–209, jun 1976.

[8] Michael L. Connolly. Shape complementarity at the hemoglobin $\alpha 1\beta 1$ subunit interface. *Biopolymers*, 25(7):1229–1247, 1986.

[9] Garrett M Morris, Ruth Huey, William Lindstrom, Michel F Sanner, Richard K Belew, David S Goodsell, and Arthur J Olson. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of computational chemistry*, 30(16):2785–2791, 2010.

[10] T J Ewing, S Makino, A G Skillman, and I D Kuntz. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des*, 15(5):411–428, 2001.

[11] Richard A. Friesner, Jay L. Banks, Robert B. Murphy, Thomas A. Halgren, Jasna J. Klicic, Daniel T. Mainz, Matthew P. Repasky, Eric H. Knoll, Mee Shelley, Jason K. Perry, David E. Shaw, Perry Francis, and Peter S. Shenkin. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*, 47(7):1739–1749, mar 2004.

[12] Peter Willett, Robert C Glen, Andrew R Leach, Robin Taylor, and Gareth Jones. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.*, 267:727–748, 1997.

[13] Gordon Lemmon, Kristian Kaufmann, and Jens Meiler. Prediction of HIV-1 Protease/Inhibitor Affinity using RosettaLigand. *Chemical Biology & Drug Design*, 79(6):888–896, jun 2012.

[14] Jie Liu and Renxiao Wang. Classification of current scoring functions. *Journal of Chemical Information and Modeling*, 55(3):475–482, 2015.

[15] Michael Berry, Burtram Fielding, and Junaid Gamieldien. Practical Considerations in Virtual Screening and Molecular Docking. In *Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology*, number AUGUST, pages 487–502. Elsevier, 2015.

[16] Ian W Davis and David Baker. RosettaLigand docking with full ligand and receptor flexibility. *J Mol Biol*, 385(2):381–392, jan 2009.

[17] J Meiler and D Baker. ROSETTALIGAND: Protein-small molecule docking with full side-chain flexibility. *Proteins*, 65(3):538–548, 2006.

[18] Samuel DeLuca, Karen Khar, and Jens Meiler. Fully Flexible Docking of Medium Sized Ligand Libraries with RosettaLigand. *Plos One*, 10(7):e0132508, 2015.

[19] Rebecca F. Alford, Andrew Leaver-Fay, Jeliazko R. Jeliazkov, Matthew J. O'Meara, Frank P. DiMaio, Hahnbeom Park, Maxim V. Shapovalov, P. Douglas Renfrew, Vikram K. Mulligan, Kalli Kappel, Jason W. Labonte, Michael S. Pacella, Richard Bonneau, Philip Bradley, Roland L. Dunbrack, Rhiju Das, David Baker, Brian Kuhlman, Tanja Kortemme, and Jeffrey J. Gray. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation*, 13(6):3031–3048, 2017.

[20] Gordon H Lemmon and Jens Meiler. *RosettaLigand docking with flexible XML protocols*, volume 819 of *Methods in Molecular Biology*. Springer New York, New York, NY, 2012.

[21] Gordon Lemmon and Jens Meiler. Towards Ligand Docking Including Explicit Interface Water Molecules. *PLoS ONE*, 8(6):e67536, jun 2013.

[22] Brittany Allison, Steven Combs, Sam DeLuca, Gordon Lemmon, Laura Mizoue, and Jens Meiler. Computational design of protein-small molecule interfaces. *Journal of Structural Biology*, 185(2):193–202, feb 2014.

[23] Ian W. Davis, Kaushik Raha, Martha S. Head, and David Baker. Blind docking of pharmaceutically relevant compounds using RosettaLigand. *Protein Science*, 18(9):1998–2002, sep 2009.

[24] Kelly L. Damm-Ganamet, Richard D. Smith, James B. Dunbar, Jeanne A. Stuckey, and Heather A. Carlson. CSAR benchmark exercise 2011-2012: evaluation of results from docking and relative ranking of blinded congeneric series. *Journal of chemical information and modeling*, 53(8):1853–70, aug 2013.

[25] Symon Gathiaka, Shuai Liu, Michael Chiu, Huanwang Yang, Jeanne A. Stuckey, You Na Kang, Jim Delproposto, Ginger Kubish, James B. Dunbar, Heather A. Carlson, Stephen K. Burley, W. Patrick Walters, Rommie E. Amaro, Victoria A. Feher, and Michael K. Gilson. D3R grand challenge 2015: Evaluation of proteinligand pose and affinity predictions. *Journal of Computer-Aided Molecular Design*, 30(9):1–18, sep 2016.

[26] Zied Gaieb, Shuai Liu, Symon Gathiaka, Michael Chiu, Huanwang Yang, Chenghua Shao, Victoria A. Feher, W. Patrick Walters, Bernd Kuhn, Markus G. Rudolph, Stephen K. Burley, Michael K. Gilson, and Rommie E. Amaro. D3R Grand Challenge 2: blind prediction of proteinligand poses, affinity rankings, and relative binding free energies. *Journal of Computer-Aided Molecular Design*, 32(1):1–20, 2018.

[27] Yan Li, Li Han, Jie Liu, Zhihai Liu, Zhixiong Zhao, Renxiao Wang, Jie Li, Li Han, Jie Liu, Zhixiong Zhao, and Renxiao Wang. Comparative Assessment of Scoring Functions on an Updated Benchmark: 1. Compilation of the Test Set. *Journal of Chemical Information and Modeling*, 54(6):1700–1716, 2014.

[28] Frank DiMaio, Yifan Song, Xueming Li, Matthias J Brunner, Chunfu Xu, Vincent Conticello, Edward Egelman, Thomas C Marlovits, Yifan Cheng, and David Baker. Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nature methods*, 12(4):361–5, 2015.

[29] Axel W. Fischer, Nathan S. Alexander, Nils Woetzel, Mert Karakas, Brian E. Weiner, and Jens Meiler. BCL::MP-fold: Membrane protein structure prediction guided by

EPR restraints. *Proteins: Structure, Function, and Bioinformatics*, 83(11):1947–1962, nov 2015.

[30] Caterina Bissantz, Bernd Kuhn, and Martin Stahl. A medicinal chemist's guide to molecular interactions. *Journal of Medicinal Chemistry*, 53(14):5061–5084, 2010.

[31] S L McGovern and B K Shoichet. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J Med Chem*, 46(14):2895–907., 2003.

[32] Peter J. Huwe, Qifang Xu, Maxim V. Shapovalov, Vivek Modi, Mark D. Andrake, and Roland L. Dunbrack. Biological function derived from predicted structures in CASP11. *Proteins: Structure, Function, and Bioinformatics*, 84(July 2015):370–391, sep 2016.

[33] Annalisa Bordogna, Alessandro Pandini, and Laura Bonati. Predicting the accuracy of protein-ligand docking on homology models. *Journal of Computational Chemistry*, 32(1):81–98, jan 2011.

[34] Philippe Ferrara and Edgar Jacoby. Evaluation of the utility of homology models in high throughput docking. *Journal of Molecular Modeling*, 13(8):897–905, 2007.

[35] Visvaldas Kairys, Miguel X. Fernandes, and Michael K. Gilson. Screening drug-like compounds by docking to homology models: A systematic study. *Journal of Chemical Information and Modeling*, 46(1):365–379, 2006.

[36] Kristian W Kaufmann and Jens Meiler. Using RosettaLigand for small molecule docking into comparative models. *PloS one*, 7(12), jan 2012.

[37] Dr Cooper. X-Ray crystallography: Assessments and validation of protein-small molecule complexes for drug discovery. *Expert Opin Drug Discov*, 6(8):771–782, 2012.

[38] Tim Ten Brink, Clémentine Aguirre, Thomas E. Exner, and Isabelle Krimm. Performance of protein-ligand docking with simulated chemical shift perturbations. *Journal of Chemical Information and Modeling*, 55(2):275–283, 2015.

[39] Ionut Onila, Tim ten Brink, Kai Fredriksson, Luca Codutti, Adam Mazur, Christian Griesinger, Teresa Carlomagno, and Thomas E. Exner. On-the-Fly Integration of Data from a Spin-Diffusion-Based NMR Experiment into ProteinLigand Docking. *Journal of Chemical Information and Modeling*, page 150814163616007, 2015.

[40] Julien Orts, Stefan Bartoschek, Christian Griesinger, Peter Monecke, and Teresa Carlomagno. An NMR-based scoring function improves the accuracy of binding pose predictions by docking by two orders of magnitude. *Journal of Biomolecular NMR*, 52(1):23–30, 2012.

[41] Lars Skjærven, Luca Codutti, Andrea Angelini, Manuela Grimaldi, Dorota Latek, Peter Monecke, Matthias K. Dreyer, and Teresa Carlomagno. Accounting for conformational variability in protein-ligand docking with nmr-guided rescoring. *Journal of the American Chemical Society*, 135(15):5819–5827, 2013.

[42] Olivier Cala, Florence Guilliere, and Isabelle Krimm. NMR-based analysis of protein-ligand interactions. *Analytical and Bioanalytical Chemistry*, 406(4):943–956, 2014.

[43] Kentaro Ishii, Masanori Noda, and Susumu Uchiyama. Mass spectrometric analysis of proteinligand interactions. *Biophysics and Physicobiology*, 13:87–95, 2016.

[44] Varnavas D. Mouchlis, Christophe Morisseau, Bruce D. Hammock, Sheng Li, J. Andrew McCammon, and Edward A. Dennis. Computer-aided drug design guided by hydrogen/deuterium exchange mass spectrometry: A powerful combination for the development of potent and selective inhibitors of Group VIA calcium-independent phospholipase A2. *Bioorganic and Medicinal Chemistry*, 24(20):4801–4811, 2016.

[45] Stephanie Leavitt and Ernesto Freire. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Current Opinion in Structural Biology*, 11(5):560–566, 2001.

[46] Ning-Ning Wei and Adel Hamza. SABRE: ligand/structure-based virtual screening approach using consensus molecular-shape pattern recognition. *Journal of chemical information and modeling*, 54(1):338–46, jan 2014.

[47] M Vieth and D J Cummins. DoMCoSAR: a novel approach for establishing the docking mode that is consistent with the structure-activity relationship. Application to HIV-1 protease inhibitors and VEGF receptor tyrosine kinase inhibitors. *Journal of medicinal chemistry*, 43(16):3020–32, aug 2000.

[48] C. J. Dinsmore, M. J. Bogusky, J. C. Culberson, J. M. Bergman, C. F. Homnick, C. B. Zartman, S. D. Mosser, M. D. Schaber, R. G. Robinson, K. S. Koblan, H. E. Huber, S. L. Graham, G. D. Hartman, J. R. Huff, and T. M. Williams. Conformational restriction of flexible ligands guided by the transferred NOE experiment: Potent macrocyclic inhibitors of farnesyltransferase [25]. *Journal of the American Chemical Society*, 123(9):2107–2108, 2001.

[49] Sheng-Yong Yang. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discovery Today*, 15(11-12):444–450, 2010.

[50] Bingjie Hu and Markus A Lill. PharmDock: a pharmacophore-based docking program. *Journal of cheminformatics*, 6(1):14, 2014.

[51] Christin Rakers, Fabian Schumacher, Walter Meinl, Hansruedi Glatt, Burkhard Kleuser, and Gerhard Wolber. In silico prediction of human sulfotransferase 1E1 activity guided by pharmacophores from molecular dynamics simulations. *Journal of Biological Chemistry*, 291(1):58–71, 2016.

[52] Dilyana Dimova, Kathrin Heikamp, Dagmar Stumpfe, and Jürgen Bajorath. Do medicinal chemists learn from activity cliffs? A systematic evaluation of cliff progression in evolving compound data sets. *Journal of Medicinal Chemistry*, 56(8):3339–3345, 2013.

[53] A. Patrícia Bento, Anna Gaulton, Anne Hersey, Louisa J. Bellis, Jon Chambers, Mark Davies, Felix A. Krüger, Yvonne Light, Lora Mak, Shaun McGlinchey, Michal Nowotka, George Papadatos, Rita Santos, and John P. Overington. The ChEMBL bioactivity database: an update. *Nucleic Acids Research*, 42(D1):D1083–D1090, jan 2014.

[54] Wallace K B Chan, Hongjiu Zhang, Jianyi Yang, Jeffrey R Brender, Junguk Hur, Arzucan Özgür, and Yang Zhang. GLASS: a comprehensive database for experimentally-validated GPCR-ligand associations. *Bioinformatics (Oxford, England)*, 31(May):btv302–, 2015.

[55] Oscar P J Van Linden, Albert J. Kooistra, Rob Leurs, Iwan J P De Esch, and Chris De Graaf. KLIFS: A knowledge-based structural database to navigate kinase-ligand interaction space. *Journal of Medicinal Chemistry*, 57(2):249–277, 2014.

[56] Huameng Li and Chenglong Li. Multiple Ligand Simultaneous Docking: Orchestrated Dancing of Ligands in Binding Sites of Protein. *Journal of computational chemistry*, 31(16):2014–2022, 2010.

[57] Timothy M Allison, Eamonn Reading, Idlir Liko, Andrew J Baldwin, Arthur Laganowsky, and Carol V Robinson. Quantifying the stabilizing effects of protein-ligand interactions in the gas phase. *Nature Communications*, 6:1–10, 2015.

[58] Andreas Evers, Holger Gohlke, and Gerhard Klebe. Ligand-supported Homology Modelling of Protein Binding-sites using Knowledge-based Potentials. *Journal of Molecular Biology*, 334(2):327–345, 2003.

[59] Elizabeth Dong Nguyen, Christoffer Norn, Thomas M Frimurer, and Jens Meiler. Assessment and Challenges of Ligand Docking into Comparative Models of G-Protein Coupled Receptors. *PLoS ONE*, 8(7):e67302, jul 2013.

[60] L C Roisman, J Piehler, J.-Y. Trosset, H A Scheraga, and G Schreiber. Structure of the interferon-receptor complex determined by distance constraints from double-mutant cycles and flexible docking. *Proceedings of the National Academy of Sciences*, 98(23):13231–13236, nov 2001.

[61] A. P. Blum, H. A. Lester, and D. A. Dougherty. Nicotinic pharmacophore: The pyridine N of nicotine and carbonyl of acetylcholine hydrogen bond across a subunit interface to a backbone NH. *Proceedings of the National Academy of Sciences*, 107(30):13206–13211, 2010.

[62] Fred Naider, Jeffrey M. Becker, Yong Hun Lee, and Amnon Horovitz. Double-mutant cycle scanning of the interaction of a peptide ligand and its G protein-coupled receptor. *Biochemistry*, 46(11):3476–3481, 2007.

[63] Catherine Marquer, Carole Fruchart-Gaillard, Guillaume Letellier, Elodie Marcon, Gilles Mourier, Sophie Zinn-Justin, André Ménez, Denis Servent, and Bernard Gilquin. Structural model of ligand-G protein-coupled receptor (GPCR) complex based on experimental double mutant cycle data: MT7 snake toxin bound to dimeric hM1 muscarinic receptor. *Journal of Biological Chemistry*, 286(36):31661–31675, 2011.

[64] Keith L. Constantine, Malcolm E. Davis, William J. Metzler, Luciano Mueller, and Brian L. Claus. Protein-ligand NOE matching: A high-throughput method for binding pose evaluation that does not require protein NMR resonance assignments. *Journal of the American Chemical Society*, 128(22):7252–7263, 2006.

[65] Julien Orts, Jennifer Tuma, Marcel Reese, S. Kaspar Grimm, Peter Monecke, Stefan

Bartoschek, Alexander Schiffer, K. Ulrich Wendt, Christian Griesinger, and Teresa Carlomagno. Crystallography-independent determination of ligand binding modes. *Angewandte Chemie - International Edition*, 47(40):7736–7740, 2008.

[66] Sheng-You Huang, Min Li, Jianxin Wang, and Yi Pan. HybridDock: A Hybrid ProteinLigand Docking Protocol Integrating Protein- and Ligand-Based Approaches. *Journal of Chemical Information and Modeling*, 56(6):1078–1087, jun 2016.

[67] Polo C.-H. Lam, Ruben Abagyan, and Maxim Totrov. Ligand-biased ensemble receptor docking (LigBEnD): a hybrid ligand/receptor structure-based approach. *Journal of Computer-Aided Molecular Design*, 32(1):187–198, jan 2018.

[68] Jeffrey Skolnick, Mu Gao, Ambrish Roy, Bharath Srinivasan, and Hongyi Zhou. Implications of the small number of distinct ligand binding pockets in proteins for drug discovery, evolution and biochemical function. *Bioorganic and Medicinal Chemistry Letters*, 25(6):1163–1170, 2015.

[69] Jonas Bostrom, Anders Hogner, and Stefan Schmitt. Do structurally similar ligands bind in a similar fashion? *Journal of medicinal chemistry*, 49(23):6716–25, nov 2006.

[70] Shipra Malhotra and John Karanicolas. When Does Chemical Elaboration Induce a Ligand To Change Its Binding Mode? *Journal of Medicinal Chemistry*, 60(1):128–145, jan 2017.

[71] Sara L. Adamski-Werner, Satheesh K. Palaninathan, James C. Sacchettini, and Jeffery W. Kelly. Diflunisal Analogues Stabilize the Native State of Transthyretin. Potent Inhibition of Amyloidogenesis. *Journal of Medicinal Chemistry*, 47(2):355–374, 2004.

[72] Hsiang Ting Lei, Zhangqi Shen, Priyanka Surana, Mathew D. Routh, Chih Chia Su,

Qijing Zhang, and Edward W. Yu. Crystal structures of CmeR-bile acid complexes from Campylobacter jejuni. *Protein Science*, 20(4):712–723, 2011.

[73] Vincent Stoll, Kent D Stewart, Clarence J Maring, Steven Muchmore, Vincent Giranda, Yu-gui Y Gu, Gary Wang, Yuanwei Chen, Minghua Sun, Chen Zhao, April L Kennedy, Darold L Madigan, Yibo Xu, Ayda Saldivar, Warren Kati, Graeme Laver, Thomas Sowin, Hing L Sham, Jonathan Greer, and Dale Kempf. Influenza neuraminidase inhibitors: structure-based design of a novel inhibitor series. *Biochemistry*, 42(3):718–27, 2003.

[74] Zhonghua Pei, Xiaofeng Li, Kenton Longenecker, Thomas W von Geldern, Paul E Wiedeman, Thomas H Lubben, Bradley A Zinker, Kent Stewart, Stephen J Ballaron, Michael a Stashko, Amanda K Mika, David W A Beno, Michelle Long, Heidi Wells, Anita J Kempf-Grote, David J Madar, Todd S McDermott, Lakshmi Bhagavatula, Michael G Fickes, Daisy Pireh, Larry R Solomon, Marc R Lake, Rohinton Edalji, Elizabeth H Fry, Hing L Sham, and James M Trevillyan. Discovery, StructureActivity Relationship, and Pharmacological Evaluation of (5-Substituted-pyrrolidinyl-2-carbonyl)-2-cyanopyrrolidines as Potent Dipeptidyl Peptidase IV Inhibitors. *Journal of Medicinal Chemistry*, 49(12):3520–3535, jun 2006.

[75] P M Fitzgerald, B M McKeever, J F VanMiddlesworth, J P Springer, J C Heimbach, C T Leu, W K Herber, R A Dixon, and P L Darke. Crystallographic analysis of a complex between human immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0-A resolution. *The Journal of biological chemistry*, 265(24):14209–14219, 1990.

[76] Ki Hwan Kim. Outliers in SAR and QSAR: 2. Is a flexible binding site a possible source of outliers? *Journal of computer-aided molecular design*, 21(8):421–35, 2007.

[77] Ki Hwan Kim. Outliers in SAR and QSAR: Is unusual binding mode a possible source of outliers? *Journal of computer-aided molecular design*, 21(8):63–86, 2007.

[78] Junichi Goto, Ryoichi Kataoka, and Noriaki Hirayama. Ph4Dock: Pharmacophore-based protein - Ligand docking. *Journal of Medicinal Chemistry*, 47(27):6804–6811, dec 2004.

[79] Manuel Rueda, Giovanni Bottegoni, and Ruben Abagyan. Consistent improvement of cross-docking results using binding site ensembles generated with elastic network normal modes. *Journal of Chemical Information and Modeling*, 49:716–725, 2009.

[80] Oliver Korb, Tim Ten Brink, Fredrick Robin Devadoss Victor Paul Raj, Matthias Keil, and Thomas E Exner. Are predefined decoy sets of ligand poses able to quantify scoring function accuracy? *Journal of computer-aided molecular design*, 26(2):185–97, mar 2012.

[81] Rommie E. Amaro, Riccardo Baron, and J. Andrew McCammon. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *Journal of Computer-Aided Molecular Design*, 22(9):693–705, 2008.

[82] Mengang Xu and Markus A. Lill. Utilizing experimental data for reducing ensemble size in flexible-protein docking. *Journal of Chemical Information and Modeling*, 52(1):187–198, 2012.

[83] Ferran Feixas, Steffen Lindert, William Sinko, and J. Andrew McCammon. Exploring the role of receptor flexibility in structure-based drug discovery. *Biophysical Chemistry*, 186:31–45, 2014.

[84] William Sinko, Steffen Lindert, and J Andrew McCammon. Accounting for receptor flexibility and enhanced sampling methods in computer-aided drug design. *Chemical biology & drug design*, 81(1):41–9, jan 2013.

[85] Andrew Anighoro, Luca Pinzi, Giulio Rastelli, and Jürgen Bajorath. Virtual Screening for Dual Hsp90/B-Raf Inhibitors. In *Methods in Pharmacology and Toxicology*, number January 2015, pages 167–188. 2017.

[86] Wenying Yu, Chenglong Li, Wenda Zhang, Yuanzheng Xia, Shanshan Li, Jia yuh Lin, Keqin Yu, Mu Liu, Lei Yang, Jianguang Luo, Yijun Chen, Hongbin Sun, and Lingyi Kong. Discovery of an Orally Selective Inhibitor of Signal Transducer and Activator of Transcription 3 Using Advanced Multiple Ligand Simultaneous Docking. *Journal of Medicinal Chemistry*, 60(7):2718–2731, 2017.

[87] Gregory Sliwoski, Sandeepkumar Kothiwale, Jens Meiler, and Edward W Lowe. Computational Methods in Drug Discovery. *Pharmacological Reviews*, 66(January):334–395, 2014.

[88] William L Jorgensen. Efficient Drug Lead Discovery and Optimizatiion. *Acc Chem Res*, 42(6):724–733, 2009.

[89] Sunghwan Kim, Paul A. Thiessen, Evan E. Bolton, Jie Chen, Gang Fu, Asta Gindulyte, Lianyi Han, Jane He, Siqian He, Benjamin A. Shoemaker, Jiyao Wang, Bo Yu, Jian Zhang, and Stephen H. Bryant. PubChem substance and compound databases. *Nucleic Acids Research*, 44(D1):D1202–D1213, 2016.

[90] Qian Liu, Brian Masek, Karl Smith, and Julian Smith. Tagged fragment method for evolutionary structure-based de novo lead generation and optimization. *Journal of medicinal chemistry*, 50(22):5392–5402, 2007.

[91] S Wang, E S Humphreys, S Y Chung, D F Delduco, S R Lustig, H Wang, K N Parker, N W Rizzo, S Subramoney, Y M Chiang, and A Jagota. Peptides with selective affinity for carbon nanotubes. *Nat Mater*, 2(3):196–200., 2003.

[92] Woody Sherman, Tyler Day, Matthew P. Jacobson, Richard A. Friesner, and Ramy

Farid. Novel procedure for modeling ligand/receptor induced fit effects. *Journal of Medicinal Chemistry*, 49(2):534–553, 2006.

[93] F Osterberg, G M Morris, M F Sanner, A J Olson, and D S Goodsell. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins*, 46(1):34–40, 2002.

[94] Sheng Y. Huang and Xiaoqin Zou. Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins: Structure, Function and Genetics*, 66:399–421, 2007.

[95] Sandeepkumar Kothiwale, Jeffrey L. Mendenhall, and Jens Meiler. BCL::Conf: Small molecule conformational sampling using a knowledge based rotamer library. *Journal of Cheminformatics*, 7(1):1–15, 2015.

[96] Marcel L. Verdonk, Paul N. Mortenson, Richard J. Hall, Michael J. Hartshorn, and Christopher W. Murray. Protein-ligand docking against non-native protein conformers. *Journal of Chemical Information and Modeling*, 48(11):2214–2225, 2008.

[97] Oliver Korb, Tjelvar S G Olsson, Simon J Bowden, Richard J Hall, Marcel L Verdonk, John W Liebeschuetz, and Jason C Cole. Potential and limitations of ensemble docking. *Journal of chemical information and modeling*, 52(5):1262–74, may 2012.

[98] Stefan G. Krimmer, Michael Betz, Andreas Heine, and Gerhard Klebe. Methyl, ethyl, propyl, butyl: Futile but not for water, as the correlation of structure and thermodynamic signature shows in a congeneric series of thermolysin inhibitors. *ChemMedChem*, 9(4):833–846, 2014.

[99] HM Ashtawy and NR Mahapatra. Does Accurate Scoring of Ligands against Protein Targets Mean Accurate Ranking? *Bioinformatics Research and Applications*, pages 298–310, 2013.

[100] Zhe Wang, Huiyong Sun, Xiaojun Yao, Dan Li, Lei Xu, Youyong Li, Sheng Tian, and Tingjun Hou. Comprehensive evaluation of ten docking programs on a diverse set of proteinligand complexes: the prediction accuracy of sampling power and scoring power. *Phys. Chem. Chem. Phys.*, 18(18):12964–12975, 2016.

[101] Heather A. Carlson, Richard D. Smith, Kelly L. Damm-Ganamet, Jeanne A. Stuckey, Aqeel Ahmed, Maire A. Convery, Donald O. Somers, Michael Kranz, Patricia A. Elkins, Guanglei Cui, Catherine E. Peishoff, Millard H. Lambert, and James B. Dunbar. CSAR 2014: A Benchmark Exercise Using Unpublished Data from Pharma. *Journal of Chemical Information and Modeling*, 56(6):1063–1077, 2016.

[102] D. E. V. Pires, T. L. Blundell, and D. B. Ascher. Platinum: a database of experimentally measured effects of mutations on structurally defined protein-ligand complexes. *Nucleic Acids Research*, 43(D1):D387–D391, 2014.

[103] Vivian Law, Craig Knox, Yannick Djoumbou, Tim Jewison, An Chi Guo, Yifeng Liu, Adam MacIejewski, David Arndt, Michael Wilson, Vanessa Neveu, Alexandra Tang, Geraldine Gabriel, Carol Ly, Sakina Adamjee, Zerihun T. Dame, Beomsoo Han, You Zhou, and David S. Wishart. DrugBank 4.0: Shedding new light on drug metabolism. *Nucleic Acids Research*, 42(D1):1091–1097, 2014.

[104] Si-Sheng Ou-Yang, Jun-Yan Lu, Xiang-Qian Kong, Zhong-Jie Liang, Cheng Luo, and Hualiang Jiang. Computational drug discovery. *Acta pharmacologica Sinica*, 33(9):1131–40, 2012.

[105] Aurélien Grosdidier, Vincent Zoete, and Olivier Michielin. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research*, 39(SUPPL. 2):270–277, 2011.

[106] S.-L. Chang and N Tjandra. Temperature dependence of protein backbone motion

from carbonyl 13C and amide 15N NMR relaxation. *J. Magn. Res.*, 174:43–53, 2005.

[107] Dina Schneidman-Duhovny, Yuval Inbar, Ruth Nussinov, and Haim J. Wolfson. PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research*, 33(SUPPL. 2):363–367, 2005.

[108] Francis Gaudreault, Louis Philippe Morency, and Rafael J. Najmanovich. NRGsuite: A PyMOL plugin to perform docking simulations in real time using FlexAID. *Bioinformatics*, 31(23):3856–3858, 2015.

[109] Zsolt Bikadi and Eszter Hazai. Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *Journal of Cheminformatics*, 1(1):1–16, 2009.

[110] Kristian W. Kaufmann, Gordon H. Lemmon, Samuel L. DeLuca, Jonathan H. Sheehan, and Jens Meiler. Practically Useful: What the R <scp>osetta</scp> Protein Modeling Suite Can Do for You. *Biochemistry*, 49(14):2987–2998, 2010.

[111] Rocco Moretti, Sergey Lyskov, Rhiju Das, Jens Meiler, and Jeffrey J. Gray. Web-accessible molecular modeling with Rosetta: The Rosetta Online Server that Includes Everyone (ROSIE). *Protein Science*, 27(1):259–268, jan 2018.

[112] Andrew Leaver-Fay, Matthew J. O'Meara, Mike Tyka, Ron Jacak, Yifan Song, Elizabeth H. Kellogg, James Thompson, Ian W. Davis, Roland A. Pache, Sergey Lyskov, Jeffrey J. Gray, Tanja Kortemme, Jane S. Richardson, James J. Havranek, Jack Snoeyink, David Baker, and Brian Kuhlman. *Scientific benchmarks for guiding macromolecular energy function improvement*, volume 523. Elsevier Inc., 1 edition, 2013.

[113] Matthew J. O'Meara, Andrew Leaver-Fay, Michael D. Tyka, Amelie Stein, Kevin Houlihan, Frank Dimaio, Philip Bradley, Tanja Kortemme, David Baker, Jack

Snoeyink, and Brian Kuhlman. Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta. *Journal of Chemical Theory and Computation*, 11(2):609–622, 2015.

[114] Marylens Hernandez, Dario Ghersi, and Roberto Sanchez. SITEHOUND-web: A server for ligand binding site identification in protein structures. *Nucleic Acids Research*, 37(SUPPL. 2):413–416, 2009.

[115] Sergey Lyskov, Fang Chieh Chou, Shane Ó. Conchúir, Bryan S. Der, Kevin Drew, Daisuke Kuroda, Jianqing Xu, Brian D. Weitzner, P. Douglas Renfrew, Parin Sripakdeevong, Benjamin Borgo, James J. Havranek, Brian Kuhlman, Tanja Kortemme, Richard Bonneau, Jeffrey J. Gray, and Rhiju Das. Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE). *PLoS ONE*, 8(5):5–7, 2013.

[116] Yifan Song, Frank DiMaio, Ray Yu-Ruei Wang, David Kim, Chris Miles, Tj Brunette, James Thompson, and David Baker. High-resolution comparative modeling with RosettaCM. *Structure (London, England : 1993)*, 21(10):1735–42, oct 2013.

[117] Steven a Combs, Samuel L DeLuca, Stephanie H DeLuca, Gordon H Lemmon, David P Nannemann, Elizabeth D Nguyen, Jordan R Willis, Jonathan H Sheehan, and Jens Meiler. Small-molecule ligand docking into comparative models with Rosetta. *Nature Protocols*, 8(7):1277–1298, jun 2013.

[118] Christian Lis, Stefan Rubner, Martin Roatsch, Angela Berg, Tyler Gilcrest, Darwin Fu, Elizabeth Nguyen, Anne-Marie Schmidt, Harald Krautscheid, Jens Meiler, and Thorsten Berg. Development of Erasin: a chromone-based STAT3 inhibitor which induces apoptosis in Erlotinib-resistant lung cancer cells. *Scientific Reports*, 7(1):17390, dec 2017.

[119] CJ Wenthur and Ryan Morrison. Discovery of (R)-(2-fluoro-4-((-4-methoxyphenyl) ethynyl) phenyl)(3-hydroxypiperidin-1-yl) methanone (ML337), an mGlu3 selective and CNS penetrant negative allosteric modulator. *Journal of medicinal . . .* , 1(56):5208–5212, 2013.

[120] W. Frank An and Nicola Tolliday. Cell-based assays for high-throughput screening. *Molecular Biotechnology*, 45(2):180–186, 2010.

[121] Joseph C. Somody, Stephen S. MacKinnon, and Andreas Windemuth. Structural coverage of the proteome for pharmaceutical applications. *Drug Discovery Today*, 22(12):1792–1799, dec 2017.

[122] Rita Santos, Oleg Ursu, Anna Gaulton, A Patrícia Bento, Ramesh S Donadi, Cristian G Bologa, Anneli Karlsson, Bissan Al-Lazikani, Anne Hersey, Tudor I Oprea, and John P Overington. A comprehensive map of molecular drug targets. *Nature Reviews Drug Discovery*, 16(1):19–34, jan 2017.

[123] Thorsten Berg. Signal Transducers and Activators of Transcription as Targets for Small Organic Molecules. *ChemBioChem*, 9(13):2039–2044, 2008.

[124] Hua Yu and Richard Jove. The STATs of cancer new molecular targets come of age. *Nature Reviews Cancer*, 4(2):97–105, 2004.

[125] Judith Müller, Bianca Sperl, Wolfgang Reindll, Anke Kiessling, and Thorsten Berg. Discovery of chromone-based inhibitors of the transcription factor STAT5. *ChemBioChem*, 9(5):723–727, 2008.

[126] Jochen Schust and Thorsten Berg. A high-throughput fluorescence polarization assay for signal transducer and activator of transcription 3. *Analytical Biochemistry*, 330(1):114–118, 2004.

[127] Alexander S. Hauser, Misty M. Attwood, Mathias Rask-Andersen, Helgi B. Schiöth, and David E. Gloriam. Trends in GPCR drug discovery: New agents, targets and indications. *Nature Reviews Drug Discovery*, 16(12):829–842, 2017.

[128] Bryan L. Roth, John J. Irwin, and Brian K. Shoichet. Discovery of new GPCR ligands to illuminate new biology. *Nature Chemical Biology*, 13(11):1143–1151, 2017.

[129] Vsevolod Katritch, Vadim Cherezov, and Raymond C. Stevens. Structure-Function of the G ProteinCoupled Receptor Superfamily. *Annual Review of Pharmacology and Toxicology*, 53(1):531–556, jan 2013.

[130] Christofer S. Tautermann. GPCR structures in drug design, emerging opportunities with new structures. *Bioorganic & Medicinal Chemistry Letters*, 24(17):4073–4079, 2014.

[131] Irina Kufareva, Manuel Rueda, and Vsevolod Katritch. Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure*, 19(8):1108–1126, 2011.

[132] Irina Kufareva, Vsevolod Katritch, RaymondC. Stevens, and Ruben Abagyan. Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges. *Structure*, 22(8):1120–1139, 2014.

[133] Jeremy Shonberg, Ralf C. Kling, Peter Gmeiner, and Stefan Löber. GPCR crystal structures: Medicinal chemistry in the pocket. *Bioorganic & Medicinal Chemistry*, 2015.

[134] P J Conn and J P Pin. Pharmacology and functions of metabotropic glutamate receptors. *Annual review of pharmacology and toxicology*, 37:205–37, jan 1997.

[135] Q J Yan, M Rammal, M Tranfaglia, and R P Bauchwitz. Suppression of two major Fragile X Syndrome mouse model phenotypes by the mGluR5 antagonist MPEP. *Neuropharmacology*, 49(7):1053–66, dec 2005.

[136] Jason M Uslaner, Sophie Parmentier-Batteur, Rosemarie B Flick, Nathaniel O Surles, June S H Lam, Caitlyn H McNaughton, Marlene a Jacobson, and Pete H Hutson. Dose-dependent effect of CDPPB, the mGluR5 positive allosteric modulator, on recognition memory is associated with GluR1 and CREB phosphorylation in the prefrontal cortex and hippocampus. *Neuropharmacology*, 57(5-6):531–8, 2009.

[137] H Awad, G W Hubert, Y Smith, A I Levey, and P J Conn. Activation of metabotropic glutamate receptor 5 has direct excitatory effects and potentiates NMDA receptor currents in neurons of the subthalamic nucleus. *J Neurosci*, 20(21):7871–7879, nov 2000.

[138] P Jeffrey Conn, Craig W Lindsley, and Carrie K Jones. Activation of metabotropic glutamate receptors as a novel approach for the treatment of schizophrenia. *Trends in pharmacological sciences*, 30(1):25–31, jan 2009.

[139] Kamondanai Hemstapat, Herve Da Costa, and Yi Nong. A novel family of potent negative allosteric modulators of group II metabotropic glutamate receptors. ... *of Pharmacology and ...*, 322(1):254–264, 2007.

[140] Filippo Caraci, Gemma Molinaro, and Giuseppe Battaglia. II Metabotropic Glutamate (mGlu) Receptors for the Treatment of Psychosis Associated with Alzheimer's Disease: Selective Activation of mGlu2 Receptors Amplifies β-. *Molecular ...*, 79(3):618–626, 2011.

[141] DJ Sheffler and CJ Wenthur. Development of a novel, CNS-penetrant, metabotropic glutamate receptor 3 (mGlu 3) NAM probe (ML289) derived from a closely related mGlu 5. *Bioorganic & medicinal ...*, 22(12):3921–3925, 2012.

[142] Colleen M CM Niswender, Kari A Johnson, Qingwei Luo, Jennifer E Ayala, Caroline Kim, P Jeffrey Conn, and C David Weaver. Gi/o-linked G protein-coupled receptor coupling to potassium channels provides new insights into the pharmacology of the group III metabotropic glutamate receptors. *Molecular pharmacology*, 73(4):1213–1224, 2008.

[143] Corrado Corti, Giuseppe Battaglia, Gemma Molinaro, Barbara Riozzi, Anna Pittaluga, Mauro Corsi, Manolo Mugnaini, Ferdinando Nicoletti, and Valeria Bruno. The use of knock-out mice unravels distinct roles for mGlu2 and mGlu3 metabotropic glutamate receptors in mechanisms of neurodegeneration/neuroprotection. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(31):8297–308, aug 2007.

[144] Andrew S. Doré, Krzysztof Okrasa, Jayesh C. Patel, Maria Serrano-Vega, Kirstie Bennett, Robert M. Cooke, James C. Errey, Ali Jazayeri, Samir Khan, Ben Tehan, Malcolm Weir, Giselle R. Wiggin, and Fiona H. Marshall. Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. *Nature*, 511:557–62, jul 2014.

[145] Huixian Wu, Karen J Gregory, P Cho, Colleen M Niswender, Jens Meiler, Vadim Cherezov, P Jeffrey Conn, and Raymond C Stevens. Structure of a Class C GPCR Metabotropic Glutamate Receptor 1 Bound to an Allosteric Modulator. *Science*, March, 2014.

[146] Gamariel Rwibasira Rudinga, Ghulam Jilany Khan, and Yi Kong. Protease-Activated Receptor 4 (PAR4): A Promising Target for Antiplatelet Therapy. *International Journal of Molecular Sciences*, 19(2):573, 2018.

[147] Justin R. Hamilton and JoAnn Trejo. Challenges and Opportunities in Protease-

Activated Receptor Drug Development. *Annual Review of Pharmacology and Toxicology*, 57(1):349–373, 2017.

[148] Cheng Zhang, Yoga Srinivasan, Daniel H. Arlow, Juan Jose Fung, Daniel Palmer, Yaowu Zheng, Hillary F. Green, Anjali Pandey, Ron O. Dror, David E. Shaw, I William, Shaun R. Coughlin, Brian K. Kobilka, William I. Weis, Shaun R. Coughlin, and Brian K. Kobilka. High-resolution crystal structure of human Protease-Activated Receptor 1 bound to the antagonist vorapaxar. *Nature*, 492(7429):387–392, 2013.

[149] Robert K. Y. Cheng, Cédric Fiez-Vandal, Oliver Schlenker, Karl Edman, Birte Aggeler, Dean G. Brown, Giles A. Brown, Robert M. Cooke, Christoph E. Dumelin, Andrew S. Doré, Stefan Geschwindner, Christoph Grebner, Nils-Olov Hermansson, Ali Jazayeri, Patrik Johansson, Louis Leong, Rudi Prihandoko, Mathieu Rappas, Holly Soutter, Arjan Snijder, Linda Sundström, Benjamin Tehan, Peter Thornton, Dawn Troast, Giselle Wiggin, Andrei Zhukov, Fiona H. Marshall, and Niek Dekker. Structural insight into allosteric modulation of protease-activated receptor 2. *Nature*, 545(7652):112–115, apr 2017.

[150] Rebecca F. Alford, Julia Koehler Leman, Brian D. Weitzner, Amanda M. Duran, Drew C. Tilley, Assaf Elazar, and Jeffrey J. Gray. An Integrated Framework Advancing Membrane Protein Modeling and Design. *PLoS Computational Biology*, 11(9):1–23, 2015.

[151] A A Canutescu, A A Shelenkov, and R L Dunbrack Jr. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci*, 12(9):2001–14., 2003.

[152] Gregory a Ross, Garrett M Morris, and Philip C Biggin. One Size Does Not Fit All: The Limits of Structure-Based Models in Drug Discovery. *Journal of chemical theory and computation*, 9(9):4266–4274, 2013.

[153] Jacob D. Durrant, Aaron J. Friedman, Kathleen E. Rogers, and J. Andrew McCammon. Comparing neural-network scoring functions and the state of the art: Applications to common library screening. *Journal of Chemical Information and Modeling*, 53(7):1726–1735, 2013.

[154] Wei Wang, Wanlin He, Xi Zhou, and Xin Chen. Optimization of molecular docking scores with support vector rank regression. *Proteins*, 81(8):1386–98, aug 2013.

[155] David Zilian and Christoph a Sotriffer. SFCscore(RF): a random forest-based scoring function for improved affinity prediction of protein-ligand complexes. *Journal of chemical information and modeling*, 53(8):1923–33, aug 2013.

[156] R S Martin, R A Henningsen, A Suen, S Apparsundaram, B Leung, Z Jia, R K Kondru, and M E Milla. Kinetic and thermodynamic assessment of binding of serotonin transporter inhibitors. *J Pharmacol Exp Ther*, 327(3):991–1000, 2008.

[157] Christian Kramer and Peter Gedeck. Leave-cluster-out cross-validation is appropriate for scoring functions derived from diverse protein data sets. *Journal of Chemical Information and Modeling*, 50(11):1961–1969, 2010.

[158] Ragul Gowthaman, Sven a Miller, Steven a Rogers, Jittasak Khowsathit, Lan Lan, Nan Bai, David K. Johnson, Chunjing Liu, Liang Xu, Asokan Anbanandam, Jeffrey Aubé, Anuradha Roy, and John Karanicolas. DARC: mapping surface topography by ray-casting for effective virtual screening at protein interaction sites. *Journal of Medicinal Chemistry*, page 150630114820003, 2015.

[159] Rona R. Ramsay, Marija R. Popovic-Nikolic, Katarina Nikolic, Elisa Uliassi, and Maria Laura Bolognesi. A perspective on multi-target drug discovery and design for complex diseases. *Clinical and Translational Medicine*, 7(1):3, 2018.

[160] Mu Gao and Jeffrey Skolnick. A Comprehensive Survey of Small-Molecule Binding Pockets in Proteins. *PLoS Computational Biology*, 9(10):e1003302, oct 2013.

[161] Connor W. Coley, Regina Barzilay, Tommi S. Jaakkola, William H. Green, and Klavs F. Jensen. Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Central Science*, 3(5):434–443, 2017.

[162] Rosa S. Kim, Nicolas Goossens, and Yujin Hoshida. Use of big data in drug development for precision medicine. *Expert Review of Precision Medicine and Drug Development*, 1(3):245–253, 2016.

[163] Ye Hu and Jürgen Bajorath. Entering the 'big data' era in medicinal chemistry: molecular promiscuity analysis revisited. *Future science OA*, 3(2):FSO179, 2017.

[164] James B. Dunbar, Richard D. Smith, Kelly L. Damm-Ganamet, Aqeel Ahmed, Emilio Xavier Esposito, James Delproposto, Krishnapriya Chinnaswamy, You-Na Na Kang, Ginger Kubish, Jason E. Gestwicki, Jeanne A. Stuckey, and Heather A. Carlson. CSAR Data Set Release 2012: Ligands, Affinities, Complexes, and Docking Decoys. *Journal of chemical information and modeling*, 50:1842–1852, may 2013.

[165] Renxiao Wang, Xueliang Fang, Yipin Lu, Chao-Yie Yang, and Shaomeng Wang. The PDBbind database: methodologies and updates. *Journal of medicinal chemistry*, 48(12):4111–9, jun 2005.

[166] Bernhard Baum, Laveena Muley, Michael Smolinski, Andreas Heine, David Hangauer, and Gerhard Klebe. Non-additivity of functional group contributions in protein-ligand binding: a comprehensive study by crystallography and isothermal titration calorimetry. *Journal of molecular biology*, 397(4):1042–54, apr 2010.

[167] Inna V. Krieger, Joel S. Freundlich, Vijay B. Gawandi, Justin P. Roberts, Vidyadhar B. Gawandi, Qingan Sun, Joshua L. Owen, Maria T. Fraile, Sofia I. Huss, Jose Luis Lavandera, Thomas R. Ioerger, and James C. Sacchettini. Structure-

Guided Discovery of Phenyl-diketo Acids as Potent Inhibitors of M. Tuberculosis Malate Synthase. *Chemistry and Biology*, 19(12):1556–1567, 2012.

[168] Michael K. Gilson, Tiqing Liu, Michael Baitaluk, George Nicola, Linda Hwang, and Jenny Chong. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Research*, 44(D1):D1045–D1053, 2016.

[169] Z. Liu, Y. Li, L. Han, J. Li, J. Liu, Z. Zhao, W. Nie, Y. Liu, and R. Wang. PDB-wide collection of binding data: current status of the PDBbind database. *Bioinformatics*, 31(3):405–412, 2014.

[170] Patrick Conway, Michael D. Tyka, Frank DiMaio, David E. Konerding, and David Baker. Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Science*, 23(1):47–55, 2014.

[171] Karen J. Gregory, Elizabeth D. Nguyen, Chrysa Malosh, Jeffrey L. Mendenhall, Jessica Z. Zic, Brittney S. Bates, Meredith J. Noetzel, Emma F. Squire, Eric M. Turner, Jerri M. Rook, Kyle a. Emmitte, Shaun R. Stauffer, Craig W. Lindsley, Jens Meiler, and P. Jeffrey Conn. Identification of specific ligand-receptor interactions that govern binding and cooperativity of diverse modulators to a common metabotropic glutamate receptor 5 allosteric site. *ACS Chemical Neuroscience*, 5(4):282–295, 2014.

[172] Sally R. Ellingson, Yinglong Miao, Jerome Baudry, and Jeremy C. Smith. Multi-conformer ensemble docking to difficult protein targets. *Journal of Physical Chemistry B*, 119(3):1026–1034, 2015.

[173] Stefano Motta and Laura Bonati. Modeling Binding with Large Conformational Changes: Key Points in Ensemble-Docking Approaches. *Journal of Chemical Information and Modeling*, 57(7):1563–1578, 2017.

[174] Andrei Zhukov, Stephen P. Andrews, James C. Errey, Nathan Robertson, Benjamin Tehan, Jonathan S. Mason, Fiona H. Marshall, Malcolm Weir, and Miles Congreve. Biophysical mapping of the adenosine A2A receptor. *Journal of medicinal chemistry*, 54(13):4312–23, jul 2011.

[175] VP Jaakola, MT Griffith, and MA Hanson. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science*, 322(5905):1211–1217, 2008.

[176] Zhenlin Yang, Shuo Han, Max Keller, Anette Kaiser, Brian J Bender, Mathias Bosse, Kerstin Burkert, Lisa M Kögler, David Wifling, Guenther Bernhardt, Nicole Plank, Timo Littmann, Peter Schmidt, Cuiying Yi, Beibei Li, Sheng Ye, Rongguang Zhang, Bo Xu, Dan Larhammar, Raymond C Stevens, Daniel Huster, Jens Meiler, Qiang Zhao, Annette G Beck-Sickinger, Armin Buschauer, and Beili Wu. Structural basis of ligand binding modes at the neuropeptide YY1 receptor. *Nature*, 556(7702):520–524, 2018.

[177] Takayuki Amemiya, Ryotaro Koike, Akinori Kidera, and Motonori Ota. PSCDB: A database for protein structural change upon ligand binding. *Nucleic Acids Research*, 40(D1):554–558, jan 2012.