MODELING PROTEIN-LIGAND INTERACTIONS

WITH APPLICATIONS TO DRUG DESIGN

A Dissertation presented to

the Faculty of the Graduate School

at the University of Missouri

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

SAM Z. GRINTER

Dr. Xiaoqin Zou, Dissertation Supervisor

MAY 2014

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

MODELING PROTEIN-LIGAND INTERACTIONS

WITH APPLICATIONS TO DRUG DESIGN

presented by Sam Z. Grinter,

a candidate for the degree of Doctor of Philosophy and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Xiaoqin Zou

Dr. Dmitry Korkin

Dr. Jianlin Cheng

Dr. Ioan Kosztin

DEDICATION

To my family, who have served as inspiration and encouragement throughout life.

ACKNOWLEDGMENTS

First and foremost, I would like to acknowledge my research advisor, Dr. Xiaoqin Zou, who has devoted a great deal of time and energy through the years to encourage my growth as a student and researcher. I would also like to acknowledge the other members of my committee, Dr. Dmitry Korkin, Dr. Jianlin Cheng, and Dr. Ioan Kosztin, each of whom has been a source of encouragement and training.

For enjoyable and enlightening discussions, I thank the members of the laboratory of Dr. Xiaoqin Zou: Dr. Sheng-You Huang, Dr. Liming Qiu, Dr. Xianjin Xu, Dr. Shan Chang, Chengfei Yan, Lin Jiang, Zhiwei Ma, and Benjamin Merideth, several of whom were also co-authors of the research in this dissertation. Additionally, I would like to acknowledge the experimentalists with whom we have collaborated during my doctoral research: those within the laboratory of Dr. Salman Hyder including Dr. Yayun Liang and those within the laboratory of Dr. Jianmin Cui. Discussions with two additional professors, Dr. Toni Kazic and Dr. Marco Ferreira, provided insights relevant to the research in this dissertation. I have worked with a few other professors on projects outside the scope of this dissertation: Dr. T. C. Huang, Dr. Linda Randall, Dr. Adam Helfer, and Dr. Peter Pfeifer, all of whom provided a fulfilling research experience.

I would like to thank the University of Missouri Informatics Institute (MUII), Department of Physics and Astronomy, Department of Mathematics, and Dr. Chi-Ren Shyu, director of MUII, for their supportive role in my undergraduate and graduate education. Support for my graduate training was also provided by the National Library of Medicine Predoctoral Fellowship (T15 LM07089, PI: Dr. Eduardo Simoes).

Finally, I would also like to thank the numerous other people who throughout my life have provided encouragement, inspiration, advice, and intellectual stimulation. There are too many to list, but I will finish by mentioning three people who were instrumental in my choice of an academic career path. Dr. Alex Iosevich provided me with a repeatedly useful background in mathematical problem-solving skills; Dr. Jerzy M. Wrobel inspired me to pursue an undergraduate degree in physics; finally, Rick L. Knowles provided me with an early exposure to biology and the scientific method.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Protein-ligand docking methods streamline the process of drug design by helping identify new chemical entities that bind to proteins. These methods provide a simplified model of protein-ligand interactions that may be used to predict the affinity and binding mode of a ligand for a protein receptor of interest. The protein receptor in question can be an important disease target, permitting docking methods to play the important role of hypothesis generation in early-stage drug discovery. The success of such methods depends greatly on the extent to which they can provide a good tradeoff between accuracy and computational efficiency, so researchers are very interested in the relative advantages of models that attempt to approximate the physical processes involved in protein-ligand binding. In this dissertation, we review the literature on protein-ligand docking methods, focusing especially on those methods used for structure-based drug design. We develop new methodology for deriving knowledge-based scoring functions, which are used in protein-ligand docking and many other applications. We present a large-scale evaluation of our docking methods, with special emphasis on the importance of accurate sampling of protein and ligand flexibility during docking. Finally we present a structure-based virtual screening study that serves as a practical application of the docking methods.

CHAPTER 1

Introduction

1.1 Knowledge-Based Scoring Functions

One commonly-used approach for modeling protein-ligand interactions is the knowledge-based (or statistical potential-based) scoring function. Knowledge-based scoring functions attempt to abstract the many physical interactions involved in protein-ligand binding into a simpler set of features [1–3]. Energies are assigned to this simpler set of features based on their relative prevalence in a training set of suitable examples, such as a set of crystal structures of native protein-ligand complexes [4]. Knowledge-based scoring functions have been used successfully in a wide variety of applications, but still face a number of open problems including the reference state problem and the sparse data problem [5], as described in the next chapter.

In this dissertation, we present our novel approach to handle the sparse data problem in knowledge-based scoring functions. The basis for this sparse data method is a Bayesian statistical model of the relationship between sparse data errors in knowledge-based scoring functions and the availability of training data. This model was used to choose the weights in a consensus approach, which combines the knowledge-based scoring function with an alternative force-field-based potential that does not rely on training data. This weighting scheme gives less weight to the knowledge-based scoring function for any atom-pair types or distances that occur rarely in the training data, thus providing a natural way to deal with the sparse data problem. We demonstrated the utility of our method by developing it as a scoring function for protein-ligand interactions called STScore. The performance of STScore

is promising. For example, STScore achieved a binding mode prediction success rate of 91% using the set of 100 complexes by Wang *et al.* [6], as described in Chapter 3. The methods developed in this dissertation for protein-ligand docking may be generalized to many other applications such as protein-protein docking and protein structure prediction.

1.2 Evaluation of Docking Methods

We performed further comparative evaluations of STScore and other docking methods using the full 2012 Community Structure-Activity Resource (CSAR) data set, including 757 compounds, the majority with known affinities, and 57 crystal structures [7]. We used the full CSAR data set for binding affinity prediction and active/inactive compound discrimination, and the subset with crystal structures to additionally evaluate the performance of the scoring functions on binding mode predictions. This study gives special attention to the problem of ligand and protein flexibility in docking, and the importance of adequate conformational sampling. Using the subset with crystal structures, we found that the binding affinity predictions are less sensitive to non-native ligand and protein conformations than the binding mode predictions. We also found the full CSAR data set to be more challenging in making binding mode predictions than the subset with structures. We developed a set of scripts for preparing the CSAR data set for docking, and these are offered freely to the academic community. This comparative evaluation is discussed in Chapter 4.

1.3 Biomedical Application

We also employed protein-ligand docking methods in an inverse docking study to search for new potential anti-cancer protein targets. In this study, we used our docking software package MDock to identify potential direct targets of PRIMA-1. PRIMA-1 is well known for its ability to restore mutant p53 protein's tumor suppressor function, leading to apoptosis in

several types of cancer cells [8]. The highest-ranked human protein of our PRIMA-1 docking results is oxidosqualene cyclase (OSC). In followup experiments that treat OSC as a possible anti-cancer target, we showed that both PRIMA-1 and Ro 48-8071, a known potent OSC inhibitor [9], significantly reduce the viability of BT-474 and T47-D breast cancer cells relative to normal mammary cell controls. In addition, like PRIMA-1, we found that Ro 48-8071 causes increased binding of p53 to DNA in BT-474 cells (which express mutant p53).

1.4 In a Nutshell

This dissertation is organized as follows. Chapter 2 reviews the literature on structure-based virtual database screening. In particular, it covers the background within the field of rational drug design. It also discusses the classic challenges facing docking methods and the recent methodological advances, as well as popular applications, methods of evaluation, and future directions. Chapter 3 presents new methodology for knowledge-based scoring functions. Aside from protein-ligand interactions, our method could be applied to protein structure prediction and protein-nucleic acid docking as well. In Chapter 4 we present our large-scale docking evaluation study, with results relevant to the work in Chapter 3. Finally, in Chapter 5 we present an application of structure-based virtual database screening.

CHAPTER 2

Literature Review

The work in this chapter will be published in a special issue of Molecules entitled "In-Silico Drug Design and In-Silico Screening."

Abstract

The docking methods used in structure-based virtual database screening offer the ability to quickly and cheaply estimate the affinity and binding mode of a ligand for the protein receptor of interest, such as a drug target. These methods can be used to enrich a database of compounds, so that more compounds that are subsequently experimentally tested are found to be pharmaceutically interesting. In addition, like all virtual screening methods used for drug design, structure-based virtual screening can focus on curated libraries of synthesizable compounds, helping to reduce the expense of subsequent experimental verification. In this review, we introduce the protein-ligand docking methods used for structure-based drug design and other biological applications. We discuss the fundamental challenges facing these methods and some of the current methodological topics of interest. We also discuss the main approaches for applying protein-ligand docking methods. We end with a discussion of the challenging aspects of evaluating or benchmarking the accuracy of docking methods for their improvement, and discuss future directions.

2.1 Introduction

In the not-so-distant past, the effects of drugs on disease were known only by empirical observation. A century of subsequent research has revealed many intricacies in the working of cellular receptors and other drug targets, and likewise the methodology of finding small molecules that bind to specific targets has become increasingly complex. This development has been marked by the realization that the interacting surfaces of cellular receptors are chemically active and often flexible, and that these properties tend to be critical to the biological effects of the small molecules, or ligands, that bind to these receptors. Within this climate of complexity, the field of rational drug design emerged to play an important role in the search for new medications [10].

An important early step in rational drug design is the identification of a biological target of interest [11]. This target of interest may be as simple as the ligand receptor of an enzyme whose over-activity is associated with disease; a compound that attenuates the enzyme's action by competitive inhibition would be pharmaceutically interesting. However, there are many other types of drug targets and a variety of chemicals that bind to them (see Table 2.1 for a list of the most frequently-targeted gene families) [12].

Rational drug design aims to use knowledge of the biological target of interest to optimize the process of finding new medications. It may be divided into two broad categories: *de novo* drug design, in which a novel compound is designed from scratch, and virtual database screening, in which computational methods are used to search through libraries of small molecules, in order to find those that are predicted to be the most likely to bind to a drug target of interest [10]. *De novo* drug design has the advantage of versatility; only the imagination and the need to synthesize the compound in question limit its conceptual possibilities. However, this advantage can also be a disadvantage. New compounds can prove difficult or expensive to synthesize, constraining the number of new compounds that

Table 2.1: The percent distribution of the gene families targeted by FDA-approved drugs as of 2005. These statistics were compiled by Overington *et al.* from FDA data in 2005 [12].

Portion of Drugs	Family of Drug Target
26.8 %	Rhodopsin-like GPCRs
13.0 %	Nuclear receptors
7.9 %	Ligand-gated ion channels
5.5 %	Voltage-gated ion channels
4.1 %	Penicillin-binding protein
3.0 %	Myeloperoxidase-like
2.7 %	Sodium: neurotransmitter symporter family
2.3 %	Type II DNA topoisomerase
$\approx$ 35 %	(other)

may be subsequently analyzed by experiment. In addition, predicting the interactions of entirely novel compounds is inherently difficult. One approach to help mitigate the synthesis problem is combinatorial chemistry, in which an easy-to-synthesize scaffold is chemically modified at various positions of attachment, by adding chemical building blocks. This approach allows for more rapid synthesis of new chemical entities, but the diversity of the new compounds may be limited in certain respects, especially chirality [13]. Virtual database screening is another approach that helps mitigate the synthesis problem. In this approach, screening usually focuses on large databases of synthesizable compounds and makes use of knowledge of known binders to the drug targets of interest or structural knowledge of the drug receptors themselves. The performance of these three approaches can be enhanced by the identification of hot spots, which refers to small regions within a druggable binding pocket that are more likely to contribute to the binding free energy [14, 15].

2.1.1 Virtual Database Screening

The main goal of rational drug design is to find a new molecule that binds to and changes the activity of a drug target, given information about that drug target or the ligands that bind to it. In virtual database screening, one uses computational techniques to search a database of

compounds for small molecules predicted to bind to a drug target [16]. Such predictions are not meant to replace experimental affinity determination, but virtual screening methods can compliment the experimental methods by producing an enriched subset of a large chemical database; the enriched subset is one in which the proportion of compounds that actually bind to the drug target of interest is increased, compared to the proportion within the whole database [17]. Thus, compounds from the subset that pass the initial virtual screening are found to be pharmaceutically interesting at a higher rate and at a lower cost.

In principal, the methods used in virtual screening may be applied to any conceivable compounds, but in practice one usually focuses on curated libraries of purchasable or synthesizable compounds, or close analogues of such compounds. Some examples include Accelrys Available Chemicals Directory (Accelrys, Inc. http://accelrys.com), eMolecules Database (eMolecules, Inc. http://www.emolecules.com/), and the free ZINC database (http://zinc.docking.org/) [18].

There are two general types of virtual screening: ligand-based virtual screening ad structure-based virtual screening. In ligand-based virtual screening, properties of a set of ligands known to bind to the drug target of interest are used to build a model for the common features believed to be important for a ligand's biological effects. This model can then be used to find new ligands that share these common features [19]. In structure-based virtual screening, the ligands are modeled as physical entities and computational techniques are used to attempt to dock the ligands to a binding sites of interest [16].

Ligand-based virtual screening

The process of optimizing the properties of a lead compound in drug design tends to employ the premise that small changes in a biologically active compound will tend to lead to small changes in its physiochemical properties [20]. The underlying premise of ligand-based screening methods is similar. In ligand-based virtual screening, the structural and chemical

properties of a ligand known to bind to a drug target of interest are used to find other compounds predicted to bind to the same target, or compounds predicted to have other similar properties which may be of interest.

There are a variety of ways that ligand-based virtual screening is undertaken. The basic physicochemical properties of molecules such as volume or partition coefficient (log P), may be related to their biological activities using regression or machine learning models such as support vector machines. The resulting models associate the biological activity of compounds to their structural properties and are therefore called quantitative structure–activity relationship models (QSAR models) [21].

Expanding on this theme, one can look at the common structural and chemical features (such hydrophobic/hydrophilic chemical groups and hydrogen bond donors/acceptors) of ligands known to bind to a drug target of interest. The analysis of these known binders can be used to build a structural map, known as a pharmacophore model, of the set of features believed to be most important for molecular recognition. One can then use this model to find other compounds that share the desired features. Producing good pharmacophore models can be difficult, due to a variety of challenges such as ambiguities in defining protonation states, but are nevertheless very popular for virtual database screening [22].

A simpler ligand-based approach is the 2D chemical similarity search. One can use the 2D connectivity data describing the active substructure of a known binder and search for analogous compounds that contain the same substructure with the database of compounds. To include more dissimilar results, one can include non-exact substructure matches that are ranked by measures of structural similarity, such as the Jaccard-Tanimoto similarity coefficient [23], or affinity fingerprints [24, 25].

Structure-based virtual screening (protein-ligand docking)

In structure-based virtual screening, a set of small molecules are computationally docked to a binding site of interest. Protein-ligand docking software attempts to sample the possible ways each ligand can be positioned in the protein receptor of interest, and provide estimates of the affinity and binding mode of a ligand for the protein receptor of interest [26–29]. Protein-ligand docking methods require a structural representation of the binding site, which may come from X-ray crystal structures, NMR experiments, or homology models [30]. The structure of the small molecules may similarly come from crystal structures, but for large-scale database screening, it is often necessary to model the possible conformations *de novo*.

There is a great variety of software packages available for performing protein-ligand docking. Some popular ones include DOCK [31–36], AutoDock [37], LUDI [38], FlexX [39, 40], GOLD [41], Glide [42, 43], and AutoDock Vina [44], in addition to MDock [45–49], developed in our laboratory. An exhaustive review of literature-cited protein-ligand docking software packages was presented in reference 50. Depending on factors such as the scoring function and sampling exhaustiveness, the docking software used to perform structure-based virtual database screening may vary greatly in speed, but is often slower than the ligand-based methods. While the ligand-based methods tend to be quite fast, they have the disadvantage that they require a set of ligands known to bind to the target. The ability of ligand-based method to find new active compounds is greatly dependent on the diversity or exhaustiveness of the set of ligands used to build the model [51]. The present review will focus primarily on structure-based methods, but will occasionally refer to ligand-based methods, given the complementary role they often play in the drug design process.

In summary, we will review the application, methodologies, and evaluation of the protein-ligand docking approaches. In Section 2.2, we will introduce the basics of computational protein-ligand docking, discussing the fundamental challenges faced by these methods, and

follow with some recent challenges that have been intensely researched in the last two years. In Section 2.3 we will sample the main approaches used in the biological application of protein-ligand docking methods. In Section 2.4 we discuss the benchmarks and evaluations used to compare the success of various protein-ligand methods. Finally, we end with some discussion and remarks about the future direction of the field.

2.2 Challenges in Protein-Ligand Docking

As aforementioned, protein-ligand docking software attempts to sample the possible ways a ligand can be positioned in a protein receptor of interest, and typically provide an estimate of the binding affinity and binding mode of a ligand for the protein receptor [26–28]. Docking involves an intrinsic trade-off between the speed of the docking algorithm and its accuracy. In an attempt to achieve higher accuracy, one can employ more advanced scoring functions or more exhaustive sampling of the possible binding modes and flexibility, but these modifications usually add to the computational cost. This tradeoff is very evident in large-scale virtual database screening, in which the number of compounds involved tends to place practical limits on the available computational time per compound. Despite all these challenges, protein-ligand docking methods have enjoyed considerable success in applications [52]. In this section, we will first introduce the basic methodologies of docking in the context of the two fundamental challenges: sampling and scoring. We will then discuss more recent methodological work.

2.2.1 Scoring in Protein-Ligand Docking

An essential component of docking methods is the scoring function. In protein-ligand docking, the scoring function typically assesses the overall favorability of a protein-ligand complex and is meant to be comparable to the free energy of binding of the protein and ligand [53, 54].

There are other attributes than one may want to score for practical reasons such as toxicity and properties related to absorption, distribution, metabolism, and excretion [55]. In addition, even within the sampling algorithm itself it may be advantageous to use more than one scoring function; for example one can use a quick, simple scoring function to discard the worst binding modes before assessing the rest more thoroughly [34, 56]. However, accurately predicting binding free energy with a general scoring function, while a lofty goal, would revolutionize the utility of the docking methods in drug design and other applications [57].

Computing the binding affinities of protein-ligand complexes cannot yet be done very accurately by a general scoring function. The calculation is especially challenging due to the combinatorial explosion of possible conformational states of the flexible protein and ligand, and of the surrounding water molecules and ions. In addition, the binding process involves a balance between many different physical interactions: a flexible ligand may gain favorable interactions upon binding, while simultaneously suffering a substantial entropic penalty as a result of binding. For example, charged polar groups may gain favorable electrostatic interactions when binding, while simultaneously losing favorable interactions with the solvent. Even when polar groups are solvent-exposed, as they like to be, this energetic favorability is partially tempered by the loss of entropic freedom of nearby water molecules. The contribution of each of these interactions can be substantial, yet tend to cancel each other out; therefore the total binding free energy involves a delicate balance, and inaccuracies in the computation of any one type of interaction can lead to substantial inaccuracies in the computation of total binding free energies [53].

Here we discuss the three main types of scoring functions used for docking. Firstly, there are the force-field-based approaches, which attempt to exhaustively model the many types of interactions involved in protein-ligand binding using physics-based functional forms and

parameters that are derived from experiments or quantum mechanical simulations. Secondly, there are empirical approaches, in which regression or machine learning methods are used to associate the desired prediction, typically the binding affinity of the complexes, with general features of those complexes such as the number of hydrogen-bonding pairs. Finally, there are statistical potentials, in which energy-like terms are assigned to structural features of protein-ligand interactions based on the frequency with which those features occur in a training set of protein-ligand complexes.

Force-field-based potentials

Force-field-based potentials can be used in protein-ligand docking as well as molecular dynamics (MD) simulations. They generally include a number of terms representing the various kinds of physical interactions that dominate protein-ligand binding. There are many popular force-field-based potentials in use for various applications, but most of them are quite similar in functional form. The main differences between them are which terms are including in the functional form and which specific values are used for the parameters in those terms. These parameters can be derived for experiment or fitted based on quantum mechanical simulations [58]. Consequently, the entities that are referred to as force fields in docking and MD simulations are typically sets of a parameters for use with the functional forms described below. One popular functional form of the force-field based potentials is the one associated with the AMBER molecular dynamics software package [59].

The AMBER force fields take the following functional form [60].

$$E_{AMBER} = E_{angle} + E_{bond} + E_{dihedral} + E_{non-bonded} \tag{2.1}$$

$$E_{angle} = \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 \tag{2.2}$$

$$E_{bond} = \sum_{\text{bonds}} K_r(l - l_0)^2 \tag{2.3}$$

$$E_{dihedral} = \sum_{\text{torsions}} \sum_{n} \frac{1}{2} V_n[1 + \cos(n\omega - \gamma)] \tag{2.4}$$

$$E_{non-bonded} = \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} f_{ij} \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} \tag{2.5}$$

E_{angle} and E_{bond} are harmonic approximations of the bond angle and strain energies, respectively, and $E_{dihedral}$ are energy terms associated with the dihedral angles of linearly-bonded sets of four atoms (especially, the backbone dihedral angels of proteins). The term $E_{non-bonded}$ aggregates the non-bonded interactions: a Lennard-Jones 6-12 potential which approximates the van der Waals attraction and Pauli replusion [61], and an electrostatic potential term.

The ff94 force field, which uses the AMBER functional form, has been very popular for simulating proteins [58], as have several subsequent versions such as AMBER 99SB force field, which differs from ff94 in the parameters associated with the backbone torsion angels [58]. The general AMBER force field (GAFF) offers parameters suitable for simulating small organic molecules such as drugs [59].

The CHARMM force fields are similar the AMBER force fields, but include some additional terms.

$$E_{CHARMM} = E_{angle} + E_{bond} + E_{dihedral} + E_{non-bonded} + E_{improper} + E_{UB} \tag{2.6}$$

$$E_{UB} = \sum_{Urey-Bradley} K_{UB}(S - S_0)^2 \tag{2.7}$$

$$E_{improper} = \sum_{impropers} K_\omega(\omega - \omega_0)^2 \tag{2.8}$$

The terms E_{bond}, E_{angle}, $E_{dihedral}$, and $E_{non-bonded}$ have functional forms like those given in equations 2.2–2.5, but the parameter values may differ. The Urey-Bradley term, E_{UB} [62], is based on the distance between the outer atoms when three atoms are linearly-bonded to

each other. The $E_{improper}$ term provides an energy penalty for improper dihedral angles and helps to control the interconversion of stereocenters. The parameters in E_{UB} and $E_{improper}$ can be optimized based on vibrational spectra [62]. The CHARMM22 force field is one of the popular ones that use the functional form defined in Equation 2.6 and is suitable for modeling proteins [62]. The more-recent CGenFF is suitable as a general force field for small molecules [63].

In addition to functional forms and parameterization, the force-field-based approaches are also distinguished by the method of simulating the solvent. The most obvious approach is simply to model all of the water molecules in the vicinity of a ligand-receptor explicitly and their interactions with the protein and ligand. There are a variety of explicit water models, which are distinguished by the number of sites used to represent the charge distribution of each water molecules: TIP3P uses 3-sites to represent the charge of the oxygen and two hydrogens, TIP4P splits the oxygen into two sites to better represent the charge distribution, and so on [64]. Due to the large number of degrees of freedom of the water molecules, simulating them explicitly is very computationally expensive.

To simulate some of the effects of the solvent while reducing the computational expense, implicit solvent approximations were introduced. These approximations generally start with the assumption that the solvent can be treated as a continuous dielectric medium with a charge distribution and a resulting electrostatic potential that obeys the Poisson-Boltzmann (PB) equation [65–68]. The PB equation can be used directly, or alternatively, further simplifications are possible. The most common of these is the generalized Born (GB) model of solvation, in which the protein and ligand atoms are modeled as spheres with a different dielectric constant than the solvent [53, 69–75]. The PB and GB models provide adequate approximations of the electrostatic effects of the solvent. In order to also include an approximation of favorable hydrophobic-hydrophobic interactions, the solvent-accessible (SA)

surface area method may be used in combination with the PB and GB models [69]. In this method, the free energy of solvation is assumed to be proportion of the surface area of the solvent accessible atoms, where the contribution of each atom depends on its type. The resulting models, PB/SA and GB/SA respectively, provide high-speed approximations for the major energetic effects of the solvent [76–81].

To further decrease the complexity, there are empirical solvent methods. In these methods, the electrostatic forces between the protein and ligand are modulated by an empirical distance-dependent parameter that roughly models the tendency of water to screen the electrostatic forces between charged atoms. For one example, this approach was used in DOCK [32].

Finally, it is worth noting that the force-field-based potentials mentioned above give estimates of the internal energy of the protein-ligand system in a specific microstate rather than the free energy of binding. In principal, one can use direct integration of the partition function to compute the free energy. Computing the full partition function is typically intractable, but sometimes approximations of the partition function are used [82]. In practice, to estimate the binding free energy using force-field-based potentials, it is necessary to either include the entropic contribution to free energy as an additional approximate term, or to employ umbrella sampling or free energy perturbation methods [83].

Empirical scoring functions

Given their relatively complex functional forms, the force-field-based approaches described in the previous subsection are computationally intensive. In order to provide a higher-speed alternative, researchers introduced empirical scoring functions [84, 85]. Like force-field-based potentials, empirical scoring functions contain terms that are based on structural features and are often inspired by physical interactions. However, empirical scoring functions differ in that the underlying functional form of these terms are simplified in an effort to capture the

favorability of an interaction without capturing the underlying physics of the interaction [6].

Empirical scoring functions combine features such as hydrophobic contacts, hydrophilic contacts, or number of hydrogen bonds, and parameterize these features as favorable or disfavorable based on regression or machine learning methods. Typically the parameters are optimized to predict the binding affinities of a set of protein-ligand complexes that are used as a training set [84, 85]. In this way, empirical scoring functions are reminiscent of the ligand-based models mentioned previously, except that instead of building a specialized model for each drug target, one uses a diverse training set in an attempt to produce a scoring function that can interpolate the binding affinities for drug targets not considered in the training set. Nevertheless, the general performance of empirical scoring functions has been limited by over-simplifications of some of the physical interactions [86]. Some examples of empirical scoring functions include LUDI [38], ChemScore [84], and X-SCORE [87].

Statistical potentials

Besides the empirical scoring functions, there is another type of scoring function that uses a simpler functional form than the ones given in Equations 2.1 and 2.6. Statistical potential-based scoring functions (also known as knowledge-based scoring functions) assign energy-like quantities to structural features, based on the frequency with which those features are found to occur in a training set of suitable examples, such as a set of protein-ligand complexes, relative to a reference state [1–3, 5, 88]. Usually, the inverse-Boltzmann equation is used to provide the relationship between the frequency of features and the energy that is assigned to those features. For protein-ligand interactions, the energy assigned to the interaction between ligand atom type i, protein atom type j, at a distance of r_k (the distance of

the kth bin), can be computed as follows [4].

$$u_{pmf,ij}(r_k) = -k_B T \ln \left[\frac{\rho_{ij}(r_k)}{\rho_{ij,\text{ref}}} \right] \tag{2.9}$$

The quantity $\rho_{ij}(r_k)/\rho_{ij,\text{ref}}$ is the relative radial density for atom pair type ij within the training set and is a function of the binned distance r_k. The density $\rho_{ij,\text{ref}}$ associated with the reference state may be computed using an ideal gas approximation or other approaches [89, 90]

The derivation of statistical potentials using the inverse-Boltzmann relation is not necessarily physically rigorous, and typically involves some false assumptions. One example is the assumption that the occurrences of features in the training set are conditionally independent of each other, given the energies associated with those features. In applications such as protein interactions and protein structure prediction, the interdependencies neglected by the derivation can manifest in the form of problems such as the excluded volume problem [3]. However, much like naïve Bayes classifiers, statistical potentials based on the inverse-Boltzmann relation have performed well in a variety of applications, regardless of the existence of dependencies within the feature set. Some more recent works have used a multibody approach that reduces this problem [91, 92].

Another open problem in the implementation of statistical potentials is the reference state problem [3, 5, 90, 93, 94]. In order to use the inverse Boltzmann relation to assign energies to features such as protein-ligand atom pair distances as in Equation 2.9, it is necessary to define a representative non-interacting state, which provides the frequencies of features one would expect to see in the training set if the features were energetically neutral. A simple ideal gas approximation may be used, and many alternatives have been proposed [89, 95–97]. Thomas *et al.* introduced an iterative method of deriving a statistical potential that helps avoid the need to define a specific reference state. In this method, which was developed for protein folding, the interaction potentials between residues were iteratively adjusted based on the

difference between the residue pair frequencies in the native structures and the residue pair frequencies in a Boltzmann-weighted ensemble of decoy conformations [2]. The extension of this approach to atomic, distance-dependent statistical potentials is non-trivial, due to the involvement of high-dimensional parameter optimization. This challenge was addressed by later methods, which have been applied to protein-ligand interactions, protein-protein interactions, and protein-RNA interactions [45, 98–100].

Finally, there is also the sparse data problem. The inverse-Boltzmann relation in Equation 2.9 maps the observed frequency of features in a training set to the energies assigned to those features; for features that occur infrequently in the training set, the deriving energies are inaccurate or undefined. Even when very large training sets are available, the problem persists due to physically disallowed states such as very close atom pair distances (*i.e.* clashes) [101, 102].

Given the many approaches that have been used to tackle the problems mentioned above and the diverse applications, there are many examples of statistical potentials. Some of the popular ones for protein-ligand interactions include DFIRE [103], DrugScore [104, 105], ITScore [45, 46], PMF-score [4], and SMoG [106].

Consensus scoring methods

None of the existing scoring approaches are perfect, and sometimes a consensus scoring approach is used to combine the advantages of various types of scoring functions in a way that provides optimal general performance. The benefit of the consensus scoring approach derives from the fact that the advantages and disadvantage of scoring functions vary from one type to another, and from the fact that the accuracy varies depending on the intended application. An early example of consensus scoring may be found in reference 107. Other examples of the consensus scoring approach include MultiScore [108], X-Score [87], and VoteDock [109]. The scoring function presented in reference 102 to deal with the sparse data

problem can also be considered a consensus approach.

2.2.2 Sampling in Protein-Ligand Docking

The other fundamental challenge facing protein-ligand docking methods is sampling. Protein-ligand binding involves changes in the relative orientation and conformation of the ligand, as well as possible conformational changes to the protein. Docking software attempts to sample these possible changes with varying degrees of exhaustiveness [110].

The simplest approach for sampling the possible ligand binding modes is rigid docking; the docking software can simply explore the six degrees of translational and rotational freedom and filter those with poor shape complementarity before final scoring. This approach was used by older versions of DOCK [111] as well as MDock [45, 46]. Ligand flexibility can still be considered by such software by pre-computing ensembles of putative ligand conformations, using software such as OMEGA (OpenEye Scientific Software http://www.eyesopen.com) [112, 113], and rigidly docking each conformation to the protein receptor of interest.

There are also docking approaches that sample the possible ligands conformations on-the-fly. One method of on-the-fly sampling is the incremental construction method, also known as the anchor-and-grow method. A rigid central portion of the ligand is placed in the binding site, and the rest of the ligand is incrementally grown from this rigid anchor, filtering out those possibilities that clash with the protein receptor during the process [34, 114]. DOCK uses this approach [34]. Similar to above, there are also fragmentation methods, in which multiple rigid fragments are placed within the binding site and the docking software attempts to link these pieces together to reconstruct plausible conformations of the target ligand. LUDI uses this approach [38].

Another approach for sampling ligand conformations is the hierarchical docking method. In this approach, low-energy conformations for each ligand are pre-computed and aligned so

that as many atoms as possible are identically-positioned. Each ensemble of pre-generated ligand conformations is organized into a hierarchy so that similar conformations are similarly positioned within the hierarchy. Then, for each possible translation and rotation of the ligand, the docking software makes use of the hierarchical data structure to simultaneously prune or filter sets of conformations that are not sterically possible for the given translation and rotation. For example, if an atom near the rigid center of the ligand is found to clash with the protein in a given rotation/translation, the method can confidently reject all of the descendent conformations in the hierarchy for that rotation/translation, because the descendants must contain the same clash, without having to sample each descendant conformation individually [115]. The Glide software package uses hierarchical filters during ligand sampling [42, 43].

In addition to these methods of sampling ligand conformations, there are methods to handle protein flexibility. One simple approach is to rigidly dock the ligands to several putative conformations of the protein, to represent some of the protein's conformational variability [47, 48, 116, 117]. Another approach, which may be used alone or in conjunction with ensemble docking is energy minimization. Minimization may be performed using Monte Carlo methods, or gradient descent minimization to help simulate some of the induced fit that occurs when a ligand binds to a protein receptor [118]. Finally, one can also attempt to explore the conformational space of critical residues of the protein, using methods analogous to the ligand methods mentioned before. For example, AutoDock4 [119] and AutoDock Vina [44] can adjust the rotatable bonds of critical residues in order to simulate protein conformation changes during binding.

2.2.3 Recent Topics

While sampling and scoring constitute the fundamental challenges in docking, much research focuses on more specialized topics. We do not attempt to exhaustively review all of the active topics, but instead sample a few topics that have received a large amount of recent attention.

Structural water

As mentioned in Section 2.2.1, water plays an important role in protein-ligand binding, often counteracting out the attractive interactions between the protein and ligand and resulting in a delicate balance of forces that is a difficult to model accurately [120]. One aspect of the solvent that has been increasingly recognized as a major actor in protein-ligand interactions is structural water. In the vicinity of the protein and ligand, molecules of water may become bound or semi-bound in certain favorable positions, stabilized by hydrogen bonds, and these structural water molecules can play a critical role in the stability of a protein-ligand interaction [121, 122]. In work by Lie *et al.*, modeling structural water molecules was found to increase the accuracy of docking simulations, up to a binding mode success rate of 67% [123]. Other work involving the inclusion of structural or bridging water molecules into high-accuracy protein-ligand docking simulations also show improvements in accuracy [124]. In another work, the ability of Rosetta (https://www.rosettacommons.org/) [125] to reproduce the binding mode of the HIV-1 protease/protease inhibitor crystal structures was investigated, and it was found that the inclusion of just a single structural water molecule in the interface was crucial for accurate prediction of an inhibitor binding pose [126]. Docking methods improved by structural water simulation have seen practical applications, such as an inverse docking application in reference 127.

Ligand promiscuity

It has been well-recognized that drugs may bind to many targets with significant affinities, and that this drug promiscuity gives rise to a complex polypharmacology with clinical relevance to the toxicity and side effects of pharmaceuticals [12, 128]. The utility of considering such ligand promiscuity early in the drug design process has already been well-recognized. A review by Taboureau *et al.* noted the regulatory recommendation that all new drug candidates

be tested for their potential to block human Ether-a-go-go Related-Gene (hERG) potassium channel, given the substantial risk of cardiotoxic side effects such as arrhythmias [129]. They considered *in silico* screening to be a useful step in identifying cardiotoxic leads before they are given larger investments.

Besides the prediction of toxicity and side effects, the tendency of ligands to bind to several sites presents another challenge: if a ligand binds tightly somewhere, even near the desired binding site, it may still not substantially affect the drug target in question. Gowthaman *et al.* point out that this is particularly important for non-traditional drug targets, such as a target within the interface of protein-protein interactions. For such targets, compounds that bind are often inadequate, if they do not bind in a sufficiently buried manner to achieve good ligand efficiency [130]. Perez-Nueno *et al.* introduced a ligand-based approach that uses shape matching to identify promiscuous ligands [131].

One the other hand, particularly for multifaceted diseases such as cancers or metabolic disorders, it may be desirable for a drug to bind to multiple targets. Peng *et al.* review the chemogenomics approaches in which a spectrum of a ligand's interactions with many drug targets is predicted by structure- or ligand-based methods. Using these approaches, one can attempt to increase those interactions within the spectrum that are desired while simultaneously reducing unwanted interactions [132].

Accurate models of the protein receptor

Docking studies often employ comparative (or homology) models of the protein target that are based on the crystal structures of homologous proteins. The methodology behind building homology models is outside the scope of this review, but here we remark on the popularity of the approach in practice. It is notable that a number of recent successful virtual screening projects used homology models of the protein receptor, and for some the best template for homology modeling was fairly divergent from the target structure, in terms of

percent sequence identity [133–139].

Nguyen *et al.* investigated the accuracy of predicting ligand binding modes in comparative models of G-protein coupled receptors. The researchers found that for the best models with template structures over 50% sequence identity, the accuracy of binding mode prediction was within 2.9 Å RMSD (root-mean-squared standard deviation) from the native experimental structure on average. In cases of low sequence similarity, it is challenging to produce a homology model with sufficient accuracy to use as a basis for virtual screening [140, 141], but percent sequence identity is not the only useful metric. It has also been suggested that choosing a template based on ligand occupancy can yield a better homology model for docking than choosing one based on percent sequence identity [142].

2.3 Protein-Ligand Docking Approaches

Having introduced structure-based drug design and the methodologies of protein-ligand docking, we will now sample the most common research approaches in which such methods have been applied.

2.3.1 Screening for New Inhibitors

Docking methods have a long and successful history of identifying new protein inhibitors and enriching compound databases in structure-based virtual screening. Here we discuss some examples of this common application.

Recently, Mahasenan *et al.* used structure-based virtual screening to identify new inhibitors of maternal embryonic leucine zipper kinase (MELK), an important kinase target known to be involved in several types of cancer. As signaling molecules, kinase targets tend to be challenging for docking methods due to their tendency to undergo major conformation changes induced by ligand binding [143]. Their three discovered inhibitors vary in affinity

from 0.37 μM to 18 μM, and may have future applications in diseases involving mis-regulation of MELK [143].

In another recent work, Heusser *et al.* performed virtual screening study of *Gloeobacter violaceus* ligand-gated ion channel (GLIC), a bacterial homolog of $GABA_A$ receptors, to search for compounds that bind to the same site as the anesthetic propofol. Among a database of commercially available compounds, 29 compounds were experimentally tested of which 16 were found to exhibit significant inhibition of GLIC relative to dimethyl sulfoxide. The active compounds were further tested on $GABA_A$ receptors. One of the compounds, like propofol, was found to inhibit both GLIC and $GABA_A$ receptors, suggesting that the GLIC receptor may be a plausible model system for $GABA_A$ receptor ligands.

In a third example, Tahir *et al.* used MODELLER (http://salilab.org/modeller) [144] to build a homology model of TNFRSF10B protein, which is believed to inhibit tumor formation [137]. In an effort to further understand this protein, they also used protein-ligand docking to screen compounds from the Mcule compound database (https://mcule.com/database) [145] for new potential inhibitors [137].

In a final example, a series of substituted heteroaromatic piperazine and piperidine derivatives were found through virtual screening based on the structure of human enterovirus 71 capsid protein VP1. The preliminary biological evaluation revealed that two of the compounds (8e and 9e) have potent activity against EV71 and Coxsackievirus A16 with low cytotoxicity [146].

2.3.2 Hybrid Approaches for Drug Design

The structure- and ligand-based methods of performing virtual database screening are not simply competing alternatives to perform the same task. They each have unique strengths and weaknesses and can therefore play a complementary role in the drug design process and other applications. Such hybrid approaches have become increasingly popular [147]. Here

are a few recent examples.

One example of a hybrid approach may be found in Ahmed *et al.* In this work, the binding profiles of the spherical C_{60} version of fullerene and its derivatives were investigated. Aside from the remarkable physico-chemical characteristics of these molecules, fullerene and its derivatives are increasingly investigated for their unique biological effects [148]. The hybrid approach used by Ahmed *et al.* included quantum-mechanical calculations, protein–ligand docking and QSAR. They used quantum-mechanical calculations to determine geometries, dipole moments, orbital energies, and other parameters of the fullerene derivatives. They used protein-ligand docking software including AutoDock Vina [44] and Schrödinger Glide [42, 43] to search for possible binding modes of the fullerene derivatives' interactions with HIV-1 protease and to identify which residues of HIV-1 protease tend to be involved in the binding. They also compared the docking scores with experimental binding affinities. Finally they used genetic algorithms to choose a suitable QSAR model predictive of the fullerene derivatives' binding activity. The most important features in the QSAR model were found to be the 3D-molecular geometry of the fullerene derivative, its number of ring systems, and its specific topology [148].

Another work that used a hybrid approach of structure- and ligand-based methods can be found in a study identifying four inhibitors of heat shock protein 90 (Hsp90), which is an important chaperone protein and anticancer drug target [149]. In this work, the researchers built a QSAR model to perform ligand-based virtual screening [149], and use a combined ligand-based/structure-based protocol to screen 1785 compounds for their predicted ability to bind to Hsp90 [150]. 80 of the predicted compounds were further evaluated by experiment and found to inhibit Hsp90 with IC50 values between 18 and 63 μM. The compounds contain possible new molecular scaffolds capable of inhibiting Hsp90 [149, 150].

The last example of the hybrid approach that we will mention here involved DNA G-

quadruplex structures, which are found in some critical positions within the genome such as near the telomeres and gene promoter regions. Unsurprisingly, they are involved in cellular aging and cancers [151]. Alcaro *et al.* used a hybrid approach to screen a database of commercially available compounds for their predicted ability to bind G-quadruplex structures. Before this work, there were already a variety of knowns binders for G-quadruplex structures. They first screened over one million compounds from the ZINC database [18] using ligand-based methods that compared the compounds in this database to the known binders using both 2D-similarity and 3D-similarity methods. The compounds which passed this first screening were then investigated using ensemble docking simulations on a few of the conformations of telomeric G-quadruplex structures that have been structurally characterized. They analyzed the compounds with the highest docking consensus score using several experimental techniques, and determined that they had found a new G-quadruplex binding moiety [151].

2.3.3 Mechanistic Studies Using Inverse Docking

Virtual database screening studies do not always start with the identification of a drug target of interest. Often, one is interested in a compound that is known to have an important biological effect, but for which the underlying molecular mechanism is unknown [152]. Consequently, rather than looking for small molecules that bind to a binding site of interest, protein-ligand docking methods may instead be used to perform the inverse search, called inverse docking [153, 154]. Inverse docking involves some additional challenges. Relative scoring of protein-ligand complexes that differ according to the protein rather than the ligand is challenging for a number of reasons. Firstly, one needs structures or models of the protein receptors to be screened, but the structures of many proteins have not been solved. This necessitates the laborious process of gathering those proteins relevant to the research in question and determining the location of the binding sites. Alternatively, one may use a curated repository

of known drug targets, such as the Potential Drug Target Database (www.dddc.ac.cn/pdtd) [155]. Secondly, proteins are often found to exist in several closely related isoforms, so the scoring function in inverse docking is challenged by the need to rank these subtle differences [156]. Thirdly, scoring functions are usually validated on benchmarks that determine their ability to accurately rank entirely different protein-ligand complexes, or many ligands against a smaller number of proteins. Benchmarks do not usually contains many examples of the same ligand docked to many different proteins, so the performance of most docking methods is more doubtful in this application. Despite these challenges, inverse docking is a popular and useful approach.

A recent application of inverse docking may be found in reference 157. Some plant-derived isoprenoids have antiparasitic effects but the relevant molecular targets of these compounds were unknown. Noting the mortality of leishmaniasis, especially in some tropical regions due to the poor availability of resources to fight drug-resistant parasites, Ogungbe and Setzer used an inverse docking approach to investigate the underlying molecular mechanism of the relevant antiparasitic isoprenoids. Specifically, they compiled the known protein targets of the drugs used to treat Leishmania and docked the isoprenoids of interest to these proteins in order to predict which of the isoprenoids may share similar targets and to offer some clues regarding their functional mechanisms [157].

2.3.4 Docking for Detailed Binding Analysis

Another useful application of protein-ligand docking methods is to analyze the physical or chemical mechanisms involved in binding. Depending on the type of scoring function used, protein-ligand docking software may provide information about the dominant interactions involved in binding, and are especially appropriate for predicting the binding position and conformation of ligands. In Docking methods provide a picture of ligand binding that is less computationally expensive than MD simulations, at the loss of information about binding

kinetics.

As noted in Section 2.3.2 on hybrid methods, protein-ligand docking is often used to further screen or refine the results of ligand-based screening, or to provide more accurate pose selections for each ligand of interest. One example is found in research by Sakkiah *et al.* These researchers used a pharmacophore model to select 85 compounds based on drug-like properties and chemical features believed to be important in inhibition of c-Src kinases, which are involved in cell proliferation and cancer. These selected compounds were further subjected to molecular docking to provide a more thorough analysis of the suitable orientations of the compounds in the c-Src active site [158]. In another work, Nguyen *et al.* identified inhibitors of severe acute respiratory syndrome 3C-like protease using structure-based virtual screening. In addition to its role in the screening, docking software was used to analyze the hydrophobic and hydrogen bond interactions of the inhibitor compounds with amino acids in the protease active site.

The binding information provided by protein-ligand docking methods are also useful for *in silico* analysis of protein mutations and their influence on ligand binding. This is an important biological application, as it can be used to help design ligands that overcome viral or bacterial drug resistances. In a study by Yang *et al.*, protein-ligand docking was used by the researchers to make predictions about the importance of neuraminidase mutations on ligand binding. The long-term goal of such research is to design new broad-spectrum antiviral drugs [159].

2.4 Docking Benchmarks and Evaluation

Benchmarking plays an important role in the development and improvement of docking methodologies [54]. Public databases combining crystal structures of protein-ligand complexes with experimentally-determined affinity data provide a standard way of assessing the

accuracy of the binding mode predictions and binding affinity predictions of protein-ligand docking methods [6, 160, 161]. In addition of the standard benchmarks, there are prospective evaluations for protein-ligand interaction predictions, also called blind competitions. These prospective evaluations play an important role in the improvement of docking methods by validating new methods on targets that were unknown to the researchers at the time the methodology was developed [7, 162, 163]. Here we discuss a number of the challenges in benchmarking docking methods.

2.4.1 Making Testable Predictions

It would be ideal for docking scoring functions, sampling schemes, and other methodologies to be tested in prospective studies in which the targets of the benchmark were unknown at the time the methodology was developed. Such prospective evaluations are not always available when new methods are introduced. In such cases, a rigorous experimental design can help ensure trustable evaluations, especially with regard to the independence of the benchmark from the development of the methods. Examples of prospective evaluations of protein-ligand docking methods include CSAR, the Community Structure Activity Resource (http://www.csardock.org) [7, 162], and OpenEye SAMPL (http://www.eyesopen.com/SAMPL) [164].

2.4.2 Assuming Lack of Knowledge of the Native, Bound Conformation

In practice, docking methods do not benefit from structurally accurate knowledge of the bound, native conformation of a protein binding site and ligand. A realistic evaluation of docking require the method to dock an arbitrary conformation of the ligand to either a ligand-free crystal structure of the protein, or if none are available, then a crystal structure bound to a different ligand than the one being docked. The allows the docking software to be tested for the available to either simulate the induced fit of binding, or test the success of a smoother scoring function designed for soft docking. Examples of recent evaluations of

docking methods that including unbound evaluations may be found in references [165] and [166].

In addition to the change in protein conformation associated with the induced fit of protein-ligand binding, docking methods also have to deal with the flexibility of ligands. To evaluate scoring function performance in flexible binding mode predictions, one approach is to pre-generate many decoy binding modes for a ligand in the vicinity of the binding site, and test the ability of the scoring function to distinguish between the native pose and decoys. This approach was used for decoy sets that extend the CSAR benchmark [161].

Korb *et al.* suggested that this approach of testing docking scoring functions with predefined sets of decoy ligands is not adequate for distinguishing a scoring function that performs well in practice from one that performs poorly. The reason is simple: in docking, sampling that is sufficient to identify the native pose and conformation must be very thorough, and so in practice many more diverse poses and conformations are considered during docking than are typically generated for the decoy poses generated for scoring function evaluation [167]. It seems that this problem could be mostly avoided by ensuring that the generated decoys are numerous and diverse, or entirely avoided by testing scoring functions simultaneously with sampling, as in realistic practice [168].

2.4.3 Assessing Binding Mode Predictions Involving Symmetric Molecules

Another challenge in evaluating docking methods is the need for special handling of symmetric molecules when evaluating binding mode predictions. The binding mode predictions of a docking method are commonly evaluated using the root-mean-squared standard deviation (RMSD) of atom positions between the known native binding mode of a ligand and its predicted mode according to the docking method. However, comparing the atom positions between two structures of a ligand requires mapping the atoms in the native ligand conformation to the atoms in the docked conformation of the ligand. Due to the symmetry

of entire molecules, or substructures within them, these mappings can be ambiguous and a naïve treatment of binding mode evaluations can consider a perfect binding mode prediction to be a poor prediction. Recently, work by Allen *et al.* addresses this problem by using the Hungarian algorithm. The Hungarian algorithm can be used to find the optimal mapping of two graphs under a cost function, and in addition to other applications in chemical informatics, has recently been used by Allen *et al.* to find the optimal mapping between two molecules in RMSD calculations, ensuring that a correct binding mode prediction will be recognized as such [169]. This method has been implemented within DOCK 6 [36] and we anticipate its wide adoption.

2.5 Discussion

The methods used to simulate the binding of proteins and small molecules face substantial challenges. Two of the fundamental challenges are sampling and scoring [26]. Protein-ligand interactions involve a delicate balance of competing forces, and these forces may occur between flexible structures that can reposition themselves in far too many combinations to sample exhaustively. Another challenge is the need to account for structural water in approximative models of the solvent [121]. In addition, there is demand for methodologies that can adequately evaluate the wide spectrum of possible binding partners of a given ligand; this is especially important for designing drugs that maximize efficacy while minimizing side effects and toxicity [12, 128]. Finally, there is also a need for rigorous, wide evaluations of new docking methodologies, which are rapidly being introduced [54, 162].

Another important component of protein-ligand binding that is sometimes neglected is the effect of entropy. Rigorous computation of the entropic contribution to binding free energy is intractable for large molecular systems such as protein-ligand complexes. Some approximations have been introduced to deal with entropy [49, 170, 171], but many of the

most popular docking scoring functions for structure-based virtual screening either ignore this important component of protein-ligand binding free energies or use overly simplistic empirical approximations. Future efforts to find computationally efficient ways to include the effect of entropy are likely to play a crucial role in the future advancement of docking methodologies.

Despite all the challenges, protein-ligand docking has a long and successful history of practical applications including newly discovered enzyme inhibitors, receptor antagonists and agonists, ion channels blockers, as well as the later approval of new drugs discovered with the help of structure-based drug design. Docking methods have provided new mechanistic insights into protein-ligand binding mechanisms, and have also helped investigate the influence of protein mutations on ligand binding, offering clues regarding the mutations that enable the robust survival of drug-resistant pathogens. As the continued increases in computational power expand the practical applications of molecular models, the field is quickly advancing, with approximations that offer a better tradeoff between accuracy and computational cost. These efforts will undoubtably lead to many more intriguing applications into the future.

CHAPTER 3

Improving Knowledge-Based Scoring Functions

The work in this chapter has been published in Journal of Computational Chemistry[*] *and was selected for the front cover of the issue.*

Abstract

Knowledge-based scoring functions are widely used for assessing putative complexes in protein-ligand and protein-protein docking and for structure prediction. Even with large training sets, knowledge-based scoring functions face the inevitable problem of sparse data. Here, we have developed a novel approach for handing the sparse data problem that is based on estimating the inaccuracies in knowledge-based scoring functions. This inaccuracy estimation is used to automatically weight the knowledge-based scoring function with an alternative, force-field-based potential (FFP) that does not rely on training data and can therefore provide an improved approximation of the interactions between rare chemical groups. The current version of STScore, a protein-ligand scoring function using our method, achieves a binding mode prediction success rate of 91% on the set of 100 complexes by Wang *et al.*, and a binding affinity correlation of 0.514 with the experimentally-determined affinities in PDBbind. The method presented here may be used with other FFPs and other knowledge-based scoring functions and can also be applied to protein-protein docking and protein structure prediction.

[*]S. Z. Grinter and X. Zou (2014) A Bayesian statistical approach of improving knowledge-based scoring functions for protein–ligand interactions. *J. Comp. Chem.* 35(12): 932–943.

3.1 Introduction

The interactions of proteins with other molecules play a fundamental role in life. Computational protein-ligand docking supports the study of such interactions by providing estimates of the mode and affinity of ligand binding, done with less expense than is possible by experimental determination [26–29]. An essential requirement of docking is the scoring function [53, 54], which assesses the favorability of a complex based on its structural features.

Knowledge-based scoring functions [2, 3, 5, 88, 172] assign energy-like quantities to feature-states, such as atom pair contacts, based on the frequency with which they are found to occur in a training set of comparable systems. This approach offers the advantage of capturing the typical energetics associated with the phenomena while using a simpler functional form than is typically possible in force-field-based approaches [53, 66–69, 88, 172, 173]. In general, a feature that is found to occur frequently relative to a reference state is considered energetically favorable. There are a variety of ways to choose a representative set of features. In protein-ligand docking, one may use contacts or other distance-dependent measures of the interactions between atoms, residues, or chemical functional groups [4, 45, 46, 49, 95, 103–105, 174–177].

In order to derive a knowledge-based scoring function from a training set of crystal structures, a common approach has been to compute a distance-dependent radial density function for the occurrences of atom pairs within a given distance of each other [4, 89, 95, 178]. For a protein-ligand scoring function, one may partition all the protein-ligand atom pairs into a set of types and then count the frequency with which the atoms of each pair type are found to occur within a given distance of each other, relative to a reference state. Based on these frequencies, the inverse-Boltzmann relationship may be used to assign energies to the interaction between ligand atom type i, protein atom type j, at a distance of r_k (the distance

of the kth bin), as follows [4, 45, 178].

$$u_{pmf,ij}(r_k) = -k_B T \ln\left[\frac{\rho_{ij}(r_k)}{\rho_{ij,\text{ref}}}\right] = -k_B T \ln\left[\frac{n_{ijk}/v_k}{N_{ij}/V}\right] \tag{3.1}$$

where $\rho_{ij}(r_k)/\rho_{ij,\text{ref}}$ is the relative radial density for atom pair type ij within the training set and is a function of the binned distance r_k. The count n_{ijk} is the number of occurrences of atom pairs of type ij found within distance bin k, and v_k is the volume of the kth distance bin. N_{ij} is the total number of atom pairs of type ij with a distance between 6.0 and 10.0 Å, and V is the total volume of a spherical shell with radii of 6.0 and 10.0 Å.

The statistical significance of $u_{pmf,ij}(r_k)$, the derived potential of mean force (PMF), is dependent on the sufficiency of the training data. In the extreme case of pair types which are absent from the training set, $u_{pmf,ij}(r_k)$ is undefined. This sparse data problem can affect any knowledge-based scoring function, because even within huge training sets there are still likely to be feature-states that never occur, such as atom pairs within unacceptably close distances. Moreover, even if an atom pair is never found within a certain distance in the training set, it may still occur in practice, such as during sampling [4].

There have been three commonly-used methods of handling the sparse data problem: the cutoff method [45, 95, 104, 176], the pseudocount method [179–181], and the method of Sippl [101]. Firstly, in the cutoff method one may simply discard features with too few instances to be trustable. For atomic pair potentials, one may discard all pair types that occur less frequently than a chosen number and treat them as non-interacting [45, 95, 104, 176]. This approach can be partially improved by combining rare types into broader groups [182], however at the expense of losing atomic specificity. The second method is pseudocounts, which are small quantities added to each count to prevent division by zero [179–181]. The added quantities do not need to be whole numbers and are flexible enough to serve as an initial guess for the atom pair densities. The most rigorous method of the three is Sippl's

method. Introduced in the context of protein structure prediction, this method treats each residue pair from the training set as a quantum of information contributing to an aggregate [101]. Starting with an initial estimate of a uniform density, Sippl's method weights the influence of each observed residue pair by a parameter so that no pair, by itself, has excessive influence on the estimated density and consequently the derived interaction potentials.

In this work, we introduce a new approach for dealing with the sparse data problem in docking. We start from the distinction between the observed counts which are obtained from the finite training set and the hypothetical exact counts [101], which can be thought of as the counts one would obtain after averaging over an infinite number of comparable training sets. The hypothetical exact counts can never be known, but we can represent their possible values as a probability density function based on the observed counts (see Methods). We then use the inverse-Boltzmann relation (equation 3.1) to derive a probability density function for the possible pair potential values given the available training data. The variance of this probability distribution provides an estimate of the error in the PMF due to the limited size of the training set. Used in conjunction with an alternative pair potential estimate, a simple force-field-based potential, we derive a weighted sum of the two that minimizes the estimated errors in the final scoring function, STScore. In this way, STScore inevitably gives more weight to the force-field-based potential when there is a lack of training data, providing an alternative way to limit the problems associated with sparse data. We conclude by showing that the performance of this consensus potential significantly exceeds both the PMF and its alternative force-field-based potential, and is competitive with other popularly-used scoring functions. The Bayesian statistical approach in this work is not limited to protein-ligand docking; it may also be applied to knowledge-based scoring functions in other applications such as protein-protein docking or protein structure prediction. This approach is also not limited to the simple force-field-based potential used in the present study; the force-field-

based potential may be replaced by other more accurate potentials or other types of scoring functions.

3.2 Methods

3.2.1 Consensus sparse data method

The problem of sparse data must be accommodated in any statistical potential. In this work we present an alternative approach to handling this problem that relies on the use of a simple force-field-based potential, u_{ffp}, as an alternative to the knowledge-based scoring function, u_{pmf}. While u_{pmf} relies on training data, u_{ffp} does not, which enables a consensus approach to handle the sparse data problem by giving more weight to the u_{ffp} for atom pair types or distances that lack sufficient examples in the training set.

For each atom pair type, we define the consensus atomic pair potential function, $u(r)$, as follows:

$$u(r) = A(r) \cdot u_{ffp}(r) + B(r) \cdot u_{pmf}(r) \tag{3.2}$$

The weights A and B are not free parameters, but mixing coefficients that are fixed by the requirement that they minimize the estimated error in u. For each atom pair type and distance, we compute the estimated random error in u_{ffp} and u_{pmf} (labeled $\mathrm{Err}(u_{ffp})$ and $\mathrm{Err}(u_{pmf})$). The detailed derivations of $\mathrm{Err}(u_{ffp})$ and $\mathrm{Err}(u_{pmf})$ are given in Methods 3.2.4 and 3.2.5. The following values of the coefficients A and B minimize $\mathrm{Err}(u)$ with the constraint that A and B sum to 1.0, as derived in the next section (Methods 3.2.2):

$$A(r) = \frac{1}{1+w} \tag{3.3}$$

$$B(r) = \frac{w}{1+w} \tag{3.4}$$

$$w = \frac{\mathrm{Err}\left(u_{\mathit{ffp}}(r)\right)}{\mathrm{Err}\left(u_{pmf}(r)\right)} \tag{3.5}$$

For each atom pair type and distance, the relative error in the two pair potentials, w, is estimated and used to determine the weighting coefficients A and B. In the extreme case of large inaccuracy in u_{pmf}, B approaches 0.0 and A approaches 1.0; namely, u_{ij} is estimated by u_{ffp} only.

Using equation 3.2, $u_{ij}(r_k)$ is computed for each protein-ligand atom pair type ij and bin k where r_k is the distance of the kth bin. The total energy score assigned to a protein-ligand complex, referred to as STScore, is then computed as the sum of the interactions of all protein-ligand atom pairs within a cutoff distance.

$$U = \sum_{\mathit{all\ L-R\ pairs\ ij}} u_{ij}(r) \tag{3.6}$$

In the present study, the cutoff distance was set to 6.0 Å.

3.2.2 Derivation of the weights A and B

For each atom pair type and distance, we wish to estimate the reliability of the two pair potentials and use the reliability estimates to choose appropriate weights A and B. To accomplish this task, we use a simplified probability model in which all the errors in u_{ffp} and u_{pmf} are considered to be random errors. $\mathrm{Err}(u_{\mathit{ffp}})$ and $\mathrm{Err}(u_{pmf})$ represent the variance of u_{ffp} and u_{pmf}, treated as random variables. The probability distribution for u_{pmf} is derived precisely in Methods 3.2.4. $\mathrm{Err}(u_{\mathit{ffp}})$ is an empirical quantity that is chosen based on physical considerations as explained in Methods 3.2.5. $\mathrm{Err}(u_{\mathit{ffp}})$ is needed as a basis of comparison for $\mathrm{Err}(u_{pmf})$.

We also use the assumption that $\mathrm{Err}(u_{\mathit{ffp}})$ is uncorrelated with $\mathrm{Err}(u_{pmf})$. This assumption is not strictly true; in some cases the errors may be positively or negatively correlated.

However, assuming that they are uncorrelated serves as a middle-point approximation and will cause no problems other than an over- or under-estimation of the ideal values for A and B in some cases. This assumption allows us to express $\mathrm{Err}(u)$ in terms of $\mathrm{Err}(u_{ffp})$ and $\mathrm{Err}(u_{pmf})$.

$$\begin{aligned} \mathrm{Err}(u) &= <(u - u_{actual})^2> & (3.7)\\ &= A^2 <(u_{ffp} - u_{actual})^2> + B^2 <(u_{pmf} - u_{actual})^2> & (3.8)\\ &= A^2 \cdot \mathrm{Err}(u_{ffp}) + B^2 \cdot \mathrm{Err}(u_{pmf}) & (3.9) \end{aligned}$$

A and B are computed separately for each atom pair type ij and distance r_k, as follows:

$$\mathrm{Err}(u_{ij}(r_k)) = A_{ijk}^2 \cdot \mathrm{Err}(u_{ffp,ij}(r_k)) + B_{ijk}^2 \cdot \mathrm{Err}(u_{pmf,ij}(r_k)) \quad (3.10)$$

As mentioned in Methods, both u_{ffp} and u_{pmf} are, by themselves, estimates of the hypothetical actual pair potential, u_{actual}, so for any choice i, j, and k, the pair of coefficients A and B should sum to 1.0:

$$A_{ij}(r_k) + B_{ij}(r_k) = 1.0 \quad (3.11)$$

Given this constraint, the function for $\mathrm{Err}(u)$ in equation 3.10 may be minimized to determine the optimal values of A and B as functions of $\mathrm{Err}(u_{ffp})$ and $\mathrm{Err}(u_{pmf})$. For example, the Lagrangian for this optimization problem (shown without indices, for simplicity) would be given by:

$$\Lambda(A, B, \lambda) = A^2\ \mathrm{Err}(u_{ffp}) + B^2\ \mathrm{Err}(u_{pmf}) + \lambda(A + B - 1) \quad (3.12)$$

If a minimum of Err(u) exists, then partial derivatives of Λ at this minimum must go to zero.

$$0 = \frac{\delta(\Lambda)}{\delta A} = 2A\ \mathrm{Err}(u_{\mathit{ffp}}) + \lambda \tag{3.13}$$

$$0 = \frac{\delta(\Lambda)}{\delta B} = 2B\ \mathrm{Err}(u_{pmf}) + \lambda \tag{3.14}$$

$$0 = \frac{\delta(\Lambda)}{\delta \lambda} = A + B - 1 \tag{3.15}$$

Choosing λ to be any scalar, one may solve for A and B as functions of Err(u_{ffp}) and Err(u_{pmf}).

$$A = \frac{1}{1 + \mathrm{Err}(u_{\mathit{ffp}})/\mathrm{Err}(u_{pmf})} \tag{3.16}$$

$$B = \frac{1}{1 + \mathrm{Err}(u_{pmf})/\mathrm{Err}(u_{\mathit{ffp}})} \tag{3.17}$$

These values of A and B exist for any pair of positive errors, Err(u_{ffp}) and Err(u_{pmf}), and minimize Err(u), to the extent possible given the assumption of uncorrelated random errors in the two constituent potentials.

We defined the quantify w in equation 3.5 as the ratio of the expected error in the two alternative pair potentials, u_{ffp} and u_{pmf}, and solved for A and B in terms of w in order to allow these weighting coefficients to be expressed more simply in equations 3.3 and 3.4 from the previous section. This notation emphasizes that only the relative accuracy of u_{ffp} and u_{pmf} influences the weights A and B, which are evaluated separately for each protein-ligand atom pair type *ij* and distance r_k.

3.2.3 Estimation of sparse data inaccuracies

The consensus method described before requires an estimate of the error in u_{pmf}. There are several potential sources of inaccuracies in u_{pmf}. In this work, we focus on estimating the contribution of sparse count data to the error, and using these estimates to minimize the influence of sparse data errors in the derived scoring function. The detailed steps are left

to the subsequent sections (Methods 3.2.4 and 3.2.5), but here we summarize the general approach.

As proposed by Sippl, it is useful to distinguish between the hypothetical exact counts (n_k^{exact} and N^{exact}) that would be obtained by averaging over an infinite number of comparable training sets, and the observed counts (n_k for the number of atom pairs observed in the kth bin, and N, the total number of atom pairs observed in the reference shell)[101]. We define u_{pmf}^{exact} as the hypothetical u_{pmf} that would be obtained if n_k^{exact} and N^{exact} were known and used in place of the observed counts n_k and N. Using Bayesian statistics, one may derive a probability density function for u_{pmf}^{exact} given the observed evidence in the training set. This probability density function represents the range of possible values for u_{pmf}^{exact}. Figure 3.1 shows three examples of this probability density function as derived for u_{pmf}^{exact}. In these three examples, the number of observed atoms in the bin is set to $n_k = 16$, $n_k = 4$, and $n_k = 1$. There are two important features in this figure. Firstly, as is typical for a u_{pmf}^{exact}, a smaller number of observed atoms in a bin corresponds to a less favorable interaction, and the distribution shifts towards the right (i.e., the positive direction). Secondly, with smaller values of n_k, the distribution becomes more spread out. As n_k decreases, the variance of the distribution, $\text{Var}(u_{pmf}^{exact})$, increases. In this way, $\text{Var}(u_{pmf}^{exact})$ provides a measure of the error due to sparse count data.

As the amount of data for a particular atom pair type and bin becomes abundant, the probability density function for u_{pmf}^{exact} approaches a Dirac delta function and $\text{Var}(u_{pmf}^{exact})$ approaches zero. However, even with non-sparse data, u_{pmf} is not a perfect representation of the actual pair potential. For this reason, we compute the estimated error between u_{pmf} and u_{actual} by adding a small constant, α, to the variance in u_{pmf}^{exact} as a rough representation of other sources of inaccuracies.

$$\text{Err}(u_{pmf}) = \text{Var}(u_{pmf}^{exact}) + \alpha \tag{3.18}$$

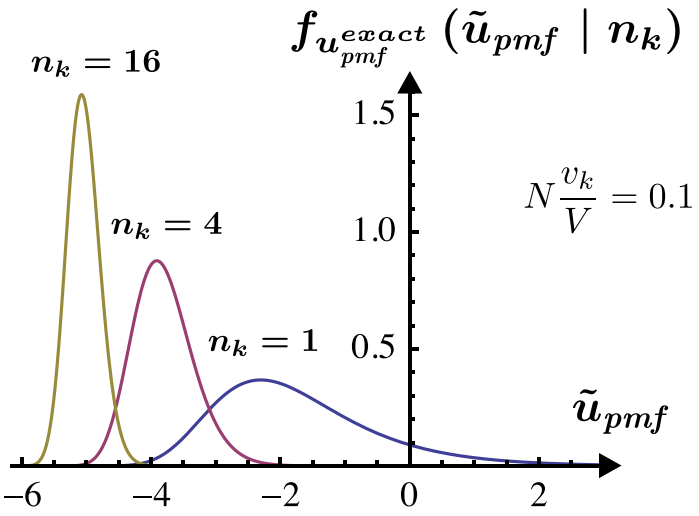

Figure 3.1: Three examples of the probability density function for u_{pmf}. The number of atom pairs within a bin, n_k, is adjusted from 1 (in blue) to 16 (in yellow) and the other parameters are held constant. As n_k increases, the variance approaches zero. In this example, the number of atoms in the reference shell is set 100, which is a typical value of N for rare atom pair types. For the pair types that are the most frequent in the training set, $N > 50000$.

Adding this constant implies that $\mathrm{Err}(u_{pmf})$ and $A(r)$ will never be computed to be exactly zero, and therefore $u_{ffp}(r)$ will always have some influence on $u(r)$, even when there is an abundance of training data. In this study, we set α to 0.1, which gave better results than other choices of α. The predictions given by STScore improve somewhat with this added value, but with or without it the method still gives better results than the PMF by itself. Methods 3.2.4 and 3.2.5 provide detailed descriptions for how we computed the error values $\mathrm{Err}(u_{pmf})$ and $\mathrm{Err}(u_{ffp})$, respectively, including a derivation of the probability density function for u_{pmf}^{exact}.

It is emphasized that the derivation of the atom-based, distance-dependent pair potentials for STScore is a one-time step. STScore is now a scoring function with tabulated potentials that is ready for use, by adding up the pairwise interactions between protein and ligand atoms (equation 3.6).

3.2.4 The potential of mean force as a probability density function

We partition the protein-ligand atom pairs into a set of types, based on the subset of the SYBYL atom types used in ref 45. Then, we count the number of times each pair type is found to occur within a given distance in a set of protein-ligand crystal structures. The energy assigned to the interaction between ligand atom type *i*, protein atom type *j*, at a distance of r_k (the distance of the *k*th bin), is based on the PMF described by Muegge in reference 178 and the subsequent work in reference 45.

$$u_{pmf,ij}(r_k) = -k_B T \ln\left[g_{ij}(r_k)\right] = -k_B T \ln\left[\frac{\rho_{ij}(r_k)}{\rho_{ij,\text{ref}}}\right] = -k_B T \ln\left[\frac{n_{ijk}/v_k}{N_{ij,\text{6-10}}/V_{\text{6-10}}}\right] \qquad (3.19)$$

The quantity $g_{ij}(r_k)$ is the radial density function for atom pair type *ij* as a function of the distance, r_k, normalized relative to the reference state. We include the $k_B T$ constant to emphasize that the computed quantity has the dimension of energy, but as is typical for a knowledge-based scoring functions, $u_{pmf,ij}(r_k)$ does not correspond to a specific physical unit and in practice $k_B T$ was set to 1.0. Here, the reference state number density, $\rho_{ij,\text{ref}}$, is determined by computing $N_{ij,\text{6-10}}$, the number of atom pairs of type *ij* with a distance between 6.0 Å and 10.0 Å, and dividing by $V_{\text{6-10}}$, the volume of a shell with these distances as radii.

We first analyze this PMF by distinguishing between the hypothetical exact counts (n_{ijk}^{exact} and N^{exact}) that would be obtained by averaging over an infinite number of comparable training sets, and the observed counts (n_{ijk} and N_{ij}), following the distinction that is presented in reference 101. The values n_{ijk}^{exact} and N_{ij}^{exact} do not need to be whole numbers. We define $u_{pmf,exact,ij}(r_k)$ as the interaction potential one would obtain from the hypothetical exact counts, for atom pair type *ij* and distance bin *k*.

$$u_{pmf,exact,ij}(r_k) = -k_B T \ln\left[\frac{\rho_{ij}(r_k)}{\rho_{ij,\text{ref}}}\right] = -k_B T \ln\left[\frac{n_{ijk}^{\text{exact}}}{N_{ij}^{\text{exact}}}\frac{V}{v_k}\right] \qquad (3.20)$$

Our aim in this subsection is to use a Bayesian statistical approach to compute a probability density function for $u_{pmf,exact,ij}$. This probability density function is computed for each protein-ligand atom pair type, *ij*, but to simplify the equations we shall hide these indices in the rest of this section. The *ij* indices should be understood to occur for the atom pair counts n_{ijk} and N_{ij} as well as the derived quantities $u_{pmf,ij}$ and $\mathrm{Err}(u_{pmf,ij})$. To find u_{pmf}^{exact}, we start with an assumption that is already implicit in the use of the inverse-Boltzmann relation (equation 3.19). That is, we assume that the probability that a pair of atoms will be observed to be within bin *k* is determined only by the potential energy associated with their interaction at a distance of r_k and the volume of the corresponding bin *k*, relative to the reference state, and conditionally independent of other considerations. In this way, when we count n_{ijk}, the number of atoms of type *ij* assigned to bin *k*, we have observed the result of a Bernoulli process, a sequence of coin tosses. However, the number of atoms of a given type is always much larger than the number assigned to a single bin, so considering it to be a Poisson process gives a very close approximation that is more mathematically convenient. We use $\tilde{n}$ to represent the possible values of n when n is unknown (that is, before a count value is observed), in contrast to the notation n_{ijk} which represents the specific count values that were observed in our training set:

$$f_n\left(\tilde{n} \mid n_{exact}\right) = \frac{\left(n_{exact}\right)^{\tilde{n}}}{\tilde{n}!} \exp[-n_{exact}] \tag{3.21}$$

Because $n << N$, we may treat N as if it were exact without any substantial underestimation of the inaccuracy in u_{pmf} due to sparse data. Including the contribution of N to the sparse data error estimate gave nearly identical results (data not shown).

Finally, we assume that the prior probability of u_{pmf}^{exact} is uniform. That is, in the absence of evidence from the training set, we consider any value for u_{pmf}^{exact} to be equally likely. This assumption allows the variance of the probability density function for u_{pmf}^{exact} to be set to

positive infinity for $n = 0$. In such cases, the weight $A(r)$ will go to 1.0 and $B(r)$ will go to 0.0, so that $u(r)$ will be determined solely by $u_{ffp}(r)$.

For values of $n > 1$, we use equations 3.20 & 3.21 to compute the probability density function for u_{pmf}^{exact}. In this equation, $\tilde{u}_{pmf}$ represents the possible values of u_{pmf}^{exact}, rather than the specific $u_{pmf,ij}(r_k)$ values that will be assigned for each atom pair type and distance at the end of this section:

$$f_{u_{pmf}^{exact}}\left(\tilde{u}_{pmf} \mid n_k\right) = \frac{1}{(n_k - 1)!}\left(N\frac{v_k}{V}\right)^{n_k} \exp\left[-\left(n_k\right)\tilde{u}_{pmf} - \left(N\frac{v_k}{V}\right)\exp\left[-\tilde{u}_{pmf}\right]\right] \tag{3.22}$$

This probability density function is a log-gamma distribution. Its variance is known to be the trigamma function, $\Psi_1(n)$:

$$\mathrm{Var}\left(u_{pmf}^{exact}\right) = \int_{-\infty}^{\infty}\left[\tilde{u}_{pmf} - \mathrm{Mean}\left(u_{pmf}^{exact}\right)\right]^2 \cdot f_{u_{pmf}^{exact}}(\tilde{u}_{pmf}) \cdot d\left(\tilde{u}_{pmf}\right) = \Psi_1(n) \tag{3.23}$$

The variance computed here is due only to stochastic variations in densities computed from sparse data. As the amount of data for a particular u_{ijk} value becomes abundant, the probability density function for u_{pmf}^{exact} approaches a Dirac delta function and its variance approaches zero (see Figure 3.1). Even with non-sparse data, the potential of mean force is not a perfect representation of the actual pair potential, so we compute the estimated error between u_{pmf} and u_{actual} by adding a small constant, 0.1, to the variance in u_{pmf}^{exact} as a rough representation of other sources of inaccuracies. Adding this empirical constant implies that $A(r)$ will never be exactly zero, and $u_{ffp}(r)$ will always have some influence on $u(r)$.

$$\mathrm{Err}(u_{pmf}) = \mathrm{Var}(u_{pmf}^{exact}) + 0.1 = \Psi_1(n) + 0.1 \tag{3.24}$$

The results of STScore are somewhat better with this empirical constant, probably due to the good performance of FFP Score.

The u_{pmf} value which is substituted into equation 3.2 may be computed by taking the most probable value from the probability density function in equation 3.22. However, this is the same value as the one computed directly from equation 3.19. With k_BT set to 1.0:

$$u_{pmf} = \text{Mode}(u_{pmf}^{exact}) = -\ln\left[\frac{n_k}{N}\frac{V}{v_k}\right] \tag{3.25}$$

Thus the statistical approach is just used to determine A and B, the mixing coefficients for u_{pmf} and u_{ffp}, as functions of the atom pair type and distance.

3.2.5 The force-field-based potential

One of the main advantages of a knowledge-based approach is the ability to get good results with a simple functional form. Moreover, our method required that the force-field-based alternative potential, u_{ffp}, also be pairwise, in order to precompute all $u_{ij}(r)$ values for computational efficiency. For these reasons we chose to use a Lennard-Jones 6-12 potential as u_{ffp} [61]. Of course, this is an unrealistically simple approximation of u_{actual}, but we consider it to be an improvement over disregarding rare atom pairs types or reducing their influence. The Lennard-Jones 6-12 well depths are based on the AMBER force field [32, 183, 184] and stored internally as kcal/mol. It was necessary to multiply these well depths by 4.2 in order for their numeric values to scale appropriately to the PMF pair potentials.

Although the focus of this work was on determining when sparse data makes u_{pmf} untrustable, STScore also requires an estimate of the inaccuracy in u_{ffp} for comparison. As a representation of the actual pair potential, the main deficiency of u_{ffp} is the lack of other kinds of interactions, such as electrostatics. We use the empirical approximation that the energies associated with these missing interactions diminish inversely with distance, r. Therefore, the variance between u_{ffp} and u_{actual}, which is the expected squared deviation between u_{ffp} and the actual pair potential, diminishes with r^2 rather than r. In other words, $\text{Err}(u_{ffp})$ can be

represented by the following function:

$$\mathrm{Err}\left(u_{ffp}\left(r\right)\right) = c\left(\frac{3.0\ \text{Å}}{r}\right)^2 \tag{3.26}$$

The parameter c represents the expected squared deviation between the u_{ffp} and the actual potential at a distance of 3.0 Å. In this work, we set c to 0.4 to correspond with the scale of the Lennard-Jones 6-12 well depths.

Now that $Err(u_{ffp})$ in equation 3.26 has been computed, it can be compared to $\mathrm{Err}(u_{pmf})$ to determine the weights A and B and compute the STScore pair potentials (see equations 3.2-3.5 of Methods 3.2.1). The distance-dependency of A and B is caused by both $\mathrm{Err}(u_{pmf})$ and $Err(u_{ffp})$, but in general $\mathrm{Err}(u_{pmf})$ is more influential because it is unbounded for close distances that do not occur in the training set. In general, u_{ffp} dominates the pair potential (i.e. $A \approx 1.0$) whenever $\mathrm{Err}(u_{pmf})$ becomes very large, and this occurs in two situations: (1) when the atom pair distance r is unusually small (i.e. clashes) and (2) when the atom pair type is rare.

3.2.6 Implementations of the other sparse data methods

Here we discuss three sparse data methods, the cutoff method, method of pseudocounts, and Sippl's method, and present the specific implementations of each which were tested in this work.

In the cutoff method one simply discards features with too few instances to be trustable. For atomic pair potentials, one may discard all pair types that occur less frequently than a chosen number and treat them as non-interacting. The specific implementation of the cutoff

method in this work is based on the sparse data method used in references 95 and 45.

$$u_{pmf,ij}^{CUTOFF}(r_k) = \begin{cases} 0 & : N_{ij,\text{0-10}} < 550 \\ u_{pmf,ij}(r_k) & : N_{ij,\text{0-10}} \geq 550 \end{cases} \tag{3.27}$$

A pair type was only regarded if 550 atom pairs of that type were found within 10.0 Å of each other in the training set, otherwise the pair type was treated as non-interacting. For unoccupied, very close distances a hard sphere interaction was used to represent the strong exchange repulsion and prevent atomic clashes.

The second sparse data approach is the method of pseudocounts. A small quantity is added to each count to prevent division by zero and to decrease the influence of sparse pair types [180]. There is flexibility in how these quantities are chosen; they may be chosen in a way that implies an initial estimate of the density distribution, which is the approach we followed in this work. For each atom pair type and distance bin, we added a pseudocount, c_k, that is proportional to the volume of the kth bin, v_k. Likewise, the pseudocount for the reference shell, $C_{\text{6-10}}$, was chosen to be proportional to the volume of the reference shell, $V_{\text{6-10}}$. The following equation is the same as equation 3.19 except for the addition of these pseudocounts.

$$u_{pmf,ij}^{PSEUDO}(r_k) = -k_B T \ln \left[\frac{\left(n_{ijk} + c_k\right)/v_k}{\left(N_{ij,\text{6-10}} + C_{\text{6-10}}\right)/V_{\text{6-10}}} \right] \tag{3.28}$$

In the absence of training data, the pair density computed for each bin in equation 3.28 would be the same, because only the pseudocounts would contribute to the count for each bin in this case. This method ensures that the derived pair potential will tend towards zero (non-interacting) wherever there is a lack of training data.

The criterion that the pseudocounts be proportional to bin volume is not sufficient to determine their exact values. It was also necessary to specify a constant representing the number of pseudocounts to add per unit volume of a bin. This density constant was chosen to

be 0.131 pseudocounts per cubic Å, implying that for atom pair types with 550 occurrences within 10 Å, about half of the counts would be from the training set and half would be pseudocounts. Therefore, this implementation of the pseudocounts method is similar to the 550-occurrence cutoff method defined previously, except that there is a gradual, rather than abrupt, transition between the treatment of sparse pairs as non-interacting and the treatment of dense pairs as trustable.

Finally there is Sippl's sparse data method [101]. Sippl's method was originally introduced as part of a knowledge-based scoring function for protein folding prediction that is based on the interactions of amino acid residues. Sippl's proposed sparse data method treats the observation of each residue in the training set as a quantum of information, contributing to an aggregate [101]. The relative influence of additional residues pairs, compared to the initial pair density estimate, are weighted by the parameter $\sigma = 0.02$. This method offers the advantage of including all the training data, while preventing sparse data noise from having excessive influence on the derived PMF. Sippl's method was introduced in the context of a statistical potential for protein folding. The energy computed with Sippl's method[101] (using the protein-ligand notation in this work) may be expressed as:

$$u^{SIPPL}_{Sippl\text{-}pmf,ij}(r_k) = k_B T \ln\left[1 + \sigma N_{ij,0\text{-}6}\right] - k_B T \ln\left[1 + \sigma N_{ij,0\text{-}6}\left(g_{ij}(r_k)\right)\right] \tag{3.29}$$

This form of the equation closely follows reference 101, but the equation may also be rearranged in terms of a weighted sum of two density estimates: the relative radial pair density computed from the training set, $g_{ij}(r_k)$, and the initial estimate of the relative radial pair density, 1.0.

$$u^{SIPPL}_{pmf,ij}(r_k) = -k_B T \ln\left[\frac{1}{1 + \sigma N_{ij,0\text{-}6}} \cdot (1.0) + \frac{\sigma N_{ij,0\text{-}6}}{1 + \sigma N_{ij,0\text{-}6}} \cdot \left(g_{ij}(r_k)\right)\right] \tag{3.30}$$

This rearrangement highlights some similarities between Sippl's method and the one presented in this work. Sippl's method uses an initial estimate of the pair density function (uniform), which is then updated after observing the training data. Similarly, STScore can be thought of as using an initial estimate of the pair energy function (u_{ffp}) which is then updated based on the training data. Dissimilarly, u_{ffp} is a non-trivial initial estimate of the pair energy function, which already gives good results in the form of FFP Score (see Results).

Sippl's method and weighting scheme may be readily applied to other PMFs by substituting the appropriate definition of $g_{ij}(r_k)$ into equation 3.30 and letting N_{ij} represent the total pool of pairs for each type that may be assigned to a distance bin. To use Sippl's method with PMF Score, we set:

$$g_{ij}(r_k) = \frac{\rho_{ij}(r_k)}{\rho_{ij,\text{ref}}} = \frac{n_{ijk}/v_k}{N_{ij,\text{6-10}}/V_{\text{6-10}}} \tag{3.31}$$

as in equation 3.19, and then computed the energy from equation 3.30.

Finally, we also evaluated an alternative to PMF Score which is based on the protein folding scoring function in Sippl's work in reference 101, except adapted for protein-ligand interactions [104]. This alternative is referred to as Sippl-PMF Score in the Results, and is the basis PMF for STScore2. In this adaptation, the relative radial pair density function is computed as:

$$g_{ij}(r_k) = \frac{n_{ijk}/N_{ij,\text{0-6}}}{n_k/N_{\text{0-6}}} \tag{3.32}$$

where $n_k/N_{\text{0-6}}$ gives the proportion of atom pairs of any type that are found in distance bin k, relative to the proportion of atom pairs within 6.0 Å. Consequently, the radial pair density function for one type is computed relative to the radial pair density for all pair types combined. This is unlike the PMF computed from equation 3.19, in which each radial pair density is computed relative to the average density for a reference shell of the same pair type.

This protein-ligand adaptation of Sippl's PMF lacks an important feature present in Sippl's

protein-folding PMF. Sippl's protein-folding PMF separates the pair density distributions into a set of topological levels; each topological level includes all the residue-residue interactions of a specified sequence distance. This separation allows the pair potential between two residues to vary not only with their geometric distance but also their sequence distance. This feature is not directly applicable to protein-ligand interactions and was not adapted.

3.2.7 Scoring function evaluations

As mentioned in Section 3.2.1, STScore is the sum of pairwise atomic interactions within a distance of 6.0 Å, using $u(r)$ from equation 3.2 as the atomic pair potential. For the purpose of evaluating STScore, we defined PMF Score and FFP Score. PMF Score is the sum of pairwise interactions within 6.0 Å using u_{pmf} from equation 3.19 as the atomic pair potential, and FFP Score likewise uses u_{ffp} for pairwire interactions within 6.0 Å. We also evaluated Sippl-PMF Score, which was defined in the same way as PMF Score except substituting the alternative relative radial pair density function given in equation 3.32 of the previous section. We evaluated two versions of STScore which differ only according to the reference state definition used for the PMF. [101] STScore1 used the PMF Score defined in equation 3.19, while STScore2 used the aforementioned Sippl-PMF Score. For simplicity in this section, we refer to the two as STScore, but all evaluations were done for both STScore1 and STScore2.

We evaluated STScore and its consensus components (PMF Score and FFP Score) based on two criteria: binding affinity predictions and binding mode predictions. For binding affinity predictions, we computed the Pearson correlation between the scores given by PMF Score, FFP Score, and STScore, and the experimentally-determined binding affinities provided in three testing sets, Wang, Muegge, and PDBbind. The testing set labeled Wang is a diverse set of 100 complexes by Wang *et al.* [6]. There is no overlap between these 100 complexes and our training set. PDBbind is a diverse set of 1300 complexes compiled by Wang *et al.* [160, 185]. There is some overlap between PDBbind and the set of training complexes used

to derive PMF Score and STScore; these duplicate complexes were removed from PDBbind, leaving 1190 testing complexes. There is no overlap between the training set and Muegge's testing set, which contains 77 complexes [95].

For binding mode predictions, we used Wang's set of 100 complexes [6] along with the 100 decoy poses generated for each complex by Morris *et al*. [37]. The heavy-atom root-mean-squared standard deviation (RMSD) was computed between each native binding pose and its respective decoy conformations. Each scoring function was used to evaluate the 100 binding modes and one native mode for each of the 100 complexes. Then, the binding modes were ranked according to their computed scores. The RMSD between the best-scored mode and the native binding mode was compared to a cutoff value, and the percentage of the 100 complexes for which the highest-scored mode was within this cutoff was computed. This percentage is a measure of the scoring function accuracy for the purpose of identifying native binding poses. The success cutoff values include 0.0, 0.5, 1.0, 1.5, and 2.0 Å.

Finally, we also evaluated PMF Score using three alternative sparse data methods: the cutoff method (the default, used for the comparisons with FFP Score) [45, 95, 104, 176], Sippl's method [101], and the method of pseudocounts [179–181], as implemented in the previous section (Methods 3.2.6).

3.3 Results and Discussion

In the Methods we described how STScore1 uses two scoring functions, PMF Score and FFP Score, as estimates of the interaction between atom pairs, and how our approach adjusts the influence of the two in order to minimize the sparse data problem. We also evaluate STScore2, which uses the same FFP with an alternative PMF that we label Sippl-PMF Score. Here, we compare the performance of STScore1, PMF Score, STScore2, Sippl-PMF Score, and FFP Score on both binding mode predictions and binding affinity predictions. We also

evaluate the performance of PMF-Score and Sippl-PMF Score when used with other sparse data methods.

First, we remark on a case study in which we used STScore1 and PMF Score to predict the binding mode of the ligand in PDB: 2cgr [186] from the Wang testing set [6, 37]. STScore1 provides a pairwise interaction for all heavy atoms pairs, regardless of how rare their type. With a cutoff of 550 instances for the inclusion of an atom-pair type, PMF Score excludes about 5% of protein-ligand atomic contacts on average based on the complexes in the Wang testing set. Using the cutoff method, PMF Score treats these excluded types as non-interacting. Figure 3.2 illustrates the contrast between the two scoring functions. In panel (a) the ligand labeled 'Native' is the binding pose from the crystal structure (PDB: 2cgr), which STScore1 correctly identified. The binding pose labeled 'PMF' is the lowest-scoring pose according to PMF Score. In this example, PMF Score's exclusion of rare atom types causes it to ignore 82 of the protein-ligand contacts (defined as protein-ligand atom pairs within 6.0 Å). Over half of these are the contacts between the protein and ligand atom N17, which is labeled and shown to be highly embedded within the binding pocket in panel (b). N17 is a rare cyano nitrogen.

We evaluated STScore1 and STScore2 along with their component potentials, and these results are shown in Figure 3.3. The left panels give the performance of STScore1 and its two component potentials, FFP Score and PMF Score. The performance of STScore2 is also shown next to its component potentials FFP Score and Sippl-PMF Score in the right panels. The top panels give the binding affinity results for each testing set. The values shown are Pearson correlation coefficients between the scores given by each scoring function and experimentally determined binding affinities for all the complexes in each testing set. Considering its simple Lennard-Jones 6-12 functional form, FFP Score (orange) gives unexpectedly good binding affinity predictions, with a Pearson correlation of 0.596 for PDBbind, 0.469 for the Wang

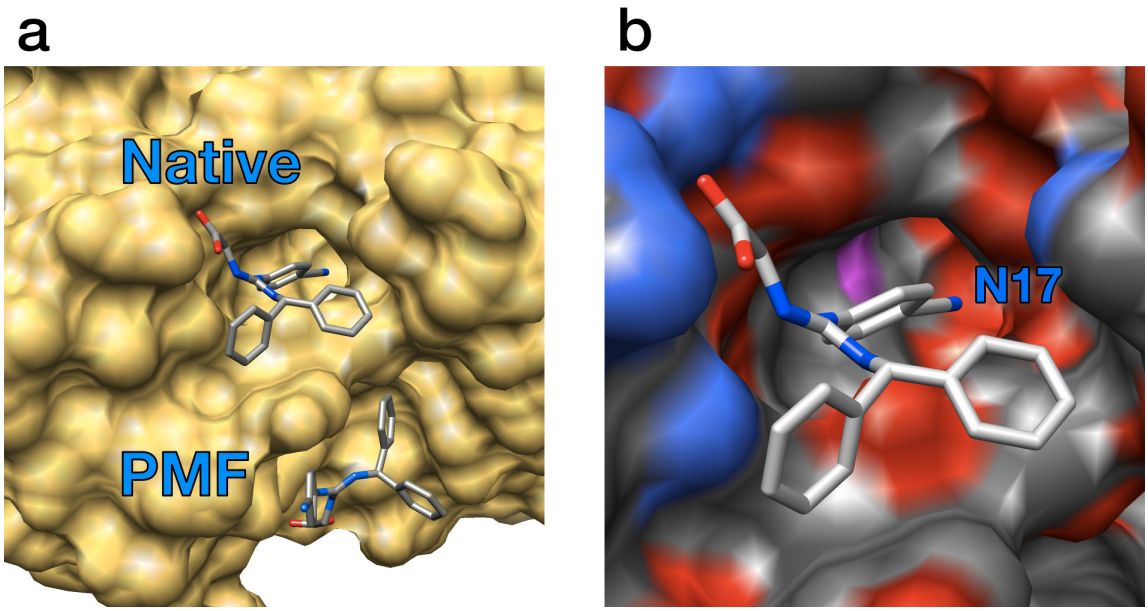

Figure 3.2: An example complex from the binding mode evaluation of STScore using Wang's testing set. In panel (a) the ligand is shown in two binding modes. The binding mode from the crystal structure (PDB: 2cgr) [186] is labeled "Native." In this example, STScore correctly identified this native binding mode by scoring it better than all of the decoy modes. The binding mode labeled "PMF" is the best-scoring mode according to PMF Score, which uses a cutoff of 550 instances in the training set for the inclusion of an atom-pair type. As a result of this cutoff, PMF Score ignores 82 of the protein-ligand contacts for this complex. Over half of these are the contacts between the protein and ligand atom N17, a relatively infrequent cyano nitrogen atom that PMF Score ignores entirely due to its rarity in the training set. N17 is highly embedded within the binding pocket in panel (b).

testing set, and 0.517 for the Muegge testing set. PMF Score (lighter blue), on the other hand, performed poorly. The Pearson correlation was -0.239 for PDBbind, -0.049 for the Wang testing set, and $+0.227$ for the Muegge testing set. STScore1 (dark grey) gave much better binding affinity predictions than PMF Score, but inconsistent improvements compared to the FFP. STScore1 was better than FFP Score on the Wang and Muegge testing sets, but worse on PDBbind. For STScore1, the Pearson correlation was 0.381 for PDBbind, 0.621 for the Wang testing set, and 0.721 for the Muegge testing set.

In the binding affinity evaluation, Sippl-PMF Score (darker blue) performed better than PMF Score. The Pearson correlation was 0.273 for PDBbind, 0.417 for the Wang testing set, and 0.366 for the Muegge testing set. Likewise STScore2 (black) gave better binding affinity predictions than Sippl-PMF Score, but inconsistent improvements compared to the FFP. STScore2 was better than FFP Score on the Wang and Muegge testing sets, but worse on PDBbind. For STScore2, the Pearson correlation was 0.514 for PDBbind, 0.607 for the Wang testing set, and 0.482 for the Muegge testing set.

We also depicted the binding affinity predictions of STScore1 and STScore2 as an array of scatterplots in Figure 3.4. The y-axis gives the scores output by STScore1 (top panels) and STScore2 (bottom panels) and the x-axis gives the known affinities according to the three testing sets, PDBbind, Wang's testing set, and Muegge's testing set. As mentioned previously, STScore1 did well in making binding affinity predictions on the Muegge testing set. Both scoring functions performed moderately well on Wang's testing set. For the large PDBbind testing set, STScore2 performed better than STScore1. We computed the best linear fit for each evaluation (shown as a red line) to indicate the approximate relationship between the docking scores and the known affinities.

Returning to Figure 3.3, the bottom panels give the binding mode prediction accuracy of STScore1, STScore2, and their component potentials for each choice of cutoff: 0.0, 0.5, 1.0,

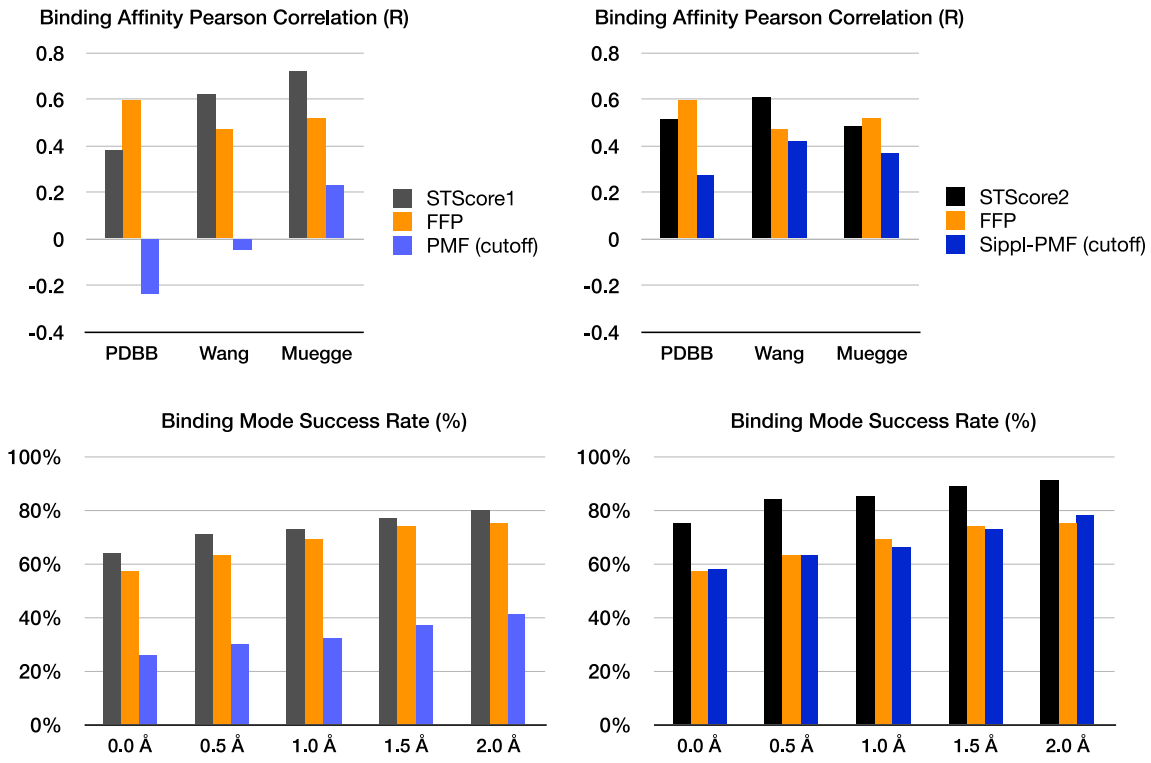

Figure 3.3: The performance of STScore1 (left panels) and STScore2 (right panels), compared to the component potentials upon which STScore1 and STScore 2 are based. STScore1 uses a force-field-based potential (FFP) and a simple potential of mean force (PMF), while STScore2 uses the same FFP along with a different potential of mean force that is based on Sippl's work (Sippl-PMF) [101]. The top panels give the binding affinity prediction accuracy as Pearson correlation coefficients between the docking scores and the experimentally-determined binding affinities in each testing set (PDBbind, Wang, and Muegge). The bottom panels give the binding mode prediction as a percent success rate, using Wang's testing set with 100 diverse protein-ligand complexes [6, 37]. A binding mode prediction is considered successful if the root-mean squared standard deviation between the highest-scoring pose and the native pose is within the given cutoff (0.0, 0.5, 1.0, 1.5, or 2.0 Å).

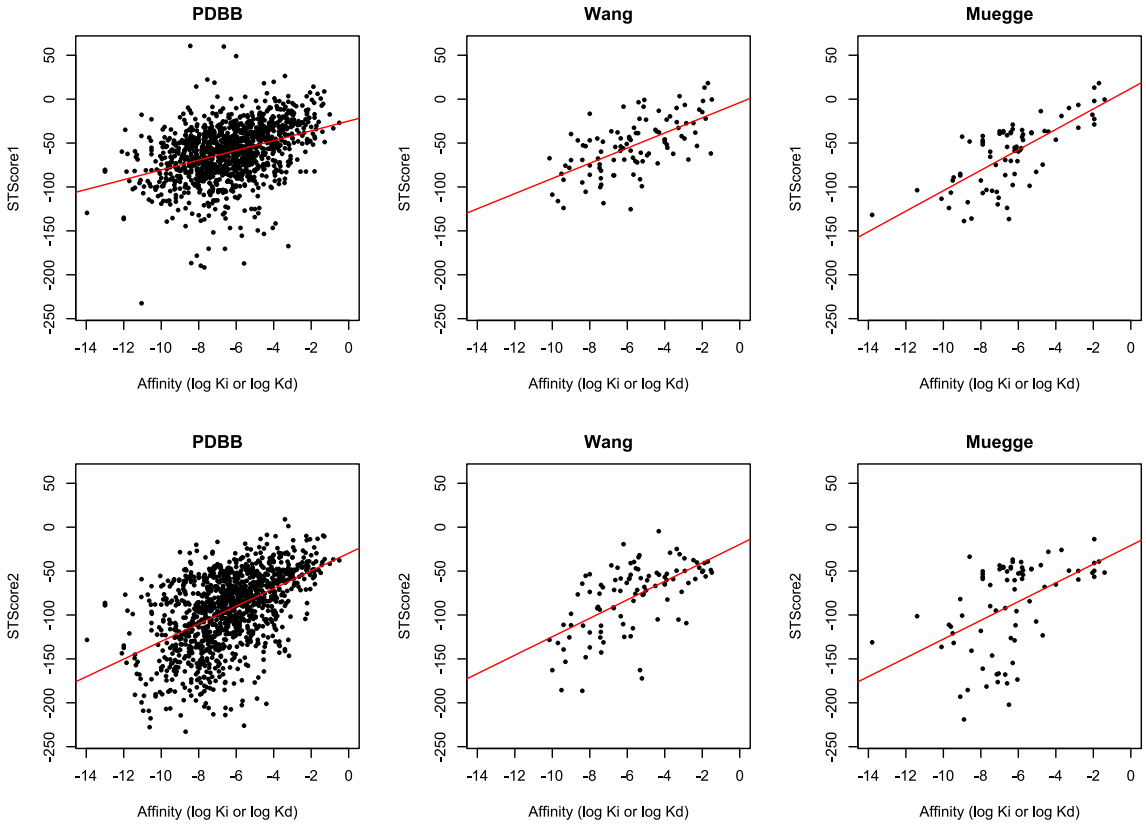

Figure 3.4: The binding affinity predictions of STScore1 and STScore2 shown as scatterplots against the known affinity values for each testing set. The *y*-axis gives the scores output by STScore1 (top panels) and STScore2 (bottom panels) and the *x*-axis gives the experimentally-determined affinities provided by each testing set. In the first column, the evaluations use the large PDBbind testing set. The second column shows the evaluations using Wang's testing set, and the third column uses Muegge's testing set. The red line shows the best linear fit for each evaluation.

1.5, and 2.0 Å. A binding mode prediction is considered successful if the root-mean-square deviation (RMSD) between the highest-scoring pose and the native pose is within the given cutoff. As shown in the figure, STScore1's binding mode predictions are consistently more accurate than PMF Score and slightly better than FFP Score. For the 2.0 Å cutoff, STScore had a binding mode prediction success rate of 80%, while FFP Score had a success rate of 75% and PMF Score a success rate of 41%. STScore2 gave good binding mode predictions that consistently exceeded the performance of both Sippl-PMF Score and FFP Score. For the 2.0 Å cutoff, STScore2 had a binding mode prediction success rate of 91%, compared to FFP Score's success rate of 75%. Sippl-PMF Score had a binding mode prediction success rate of 78%, much better than PMF Score.

The results in Figure 3.3 show that STScore1 (dark grey) and STScore2 (black) maintain the good binding affinity predictions of FFP Score while improving on both the PMF and FFP in terms of binding mode predictions. To put the performance of STScore in a relative context, we also show these results compared to other commonly-used scoring functions that have been evaluated on Wang's testing set [6, 37] in Figure 3.5.

We compared STScore to other commonly-used methods of dealing with the sparse data problem. Specifically, we tried the fixed-cutoff method as a baseline ("Cutoff"), the method of Sippl ("Sippl's method") [101], and the pseudocount method ("Pseudocounts") [180]. These three methods were adapted for and tested using the PMF defined in equations 3.1 and 3.19, and are based on references 45 and 178. Figure 3.6 compares the binding mode prediction success rate of STScore1 to PMF Score when using the three alternative sparse data methods. We also compared STScore2 to Sippl-PMF Score using the same three sparse data methods. In general, the cutoff method, pseudocounts method, and Sippl's method gave similar results.

The top panel of Figure 3.6 gives the binding affinity results for each scoring function. The y-axis is the Pearson correlation between the docking scores and known binding affinities in

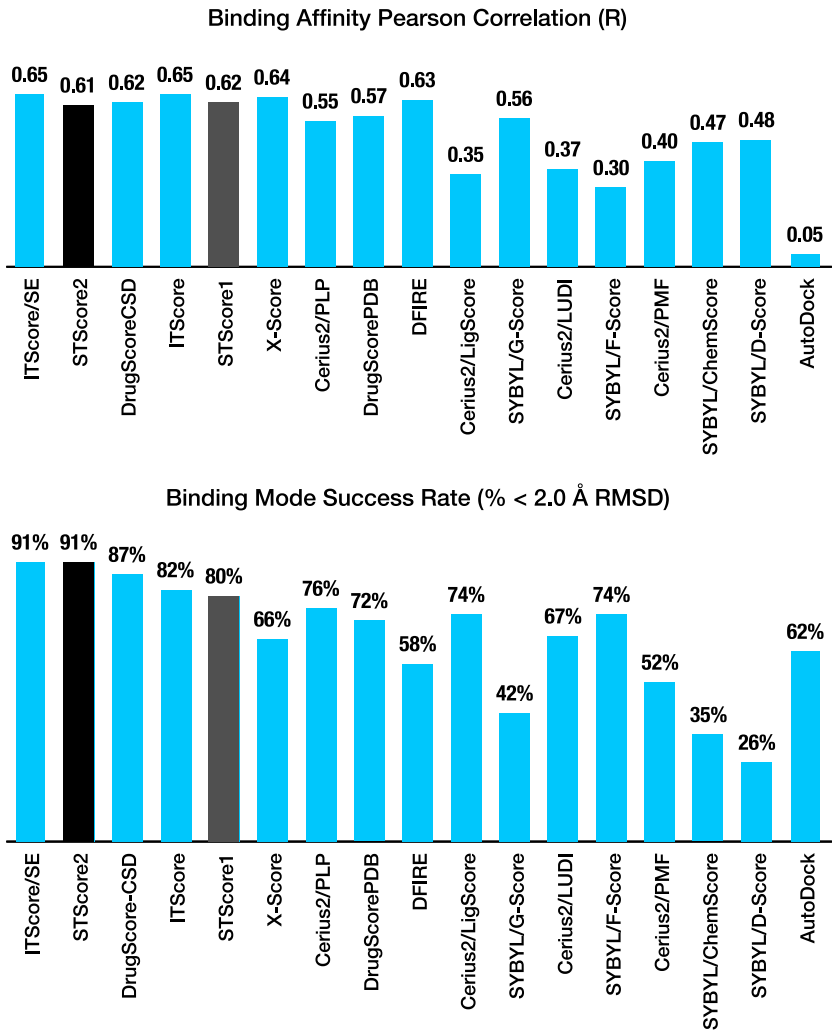

Figure 3.5: The performance of several popular scoring functions along with STScore1 and STScore2 when using the test set by Wang *et al.* with 100 diverse protein-ligand complexes [6, 37]. The top panel gives the binding affinity prediction accuracy as Pearson correlation coefficients and the bottom panel gives the binding mode prediction success rates of each scoring function. A prediction is considered successful if the best-scoring binding mode is within 2.0 Å RMSD of the native mode. Results for scoring functions other than STScore are from references 105, 103, 46, 49, and 6. The ordering of the scoring functions in this figure was chosen by normalizing the binding affinity correlations and binding mode success rates such that the highest value on each panel would be 1.0, and then averaging the two normalized performance values. We used this ordering for consistency between the top and bottom panels, and it is not meant as a general performance ranking of the listed scoring functions.

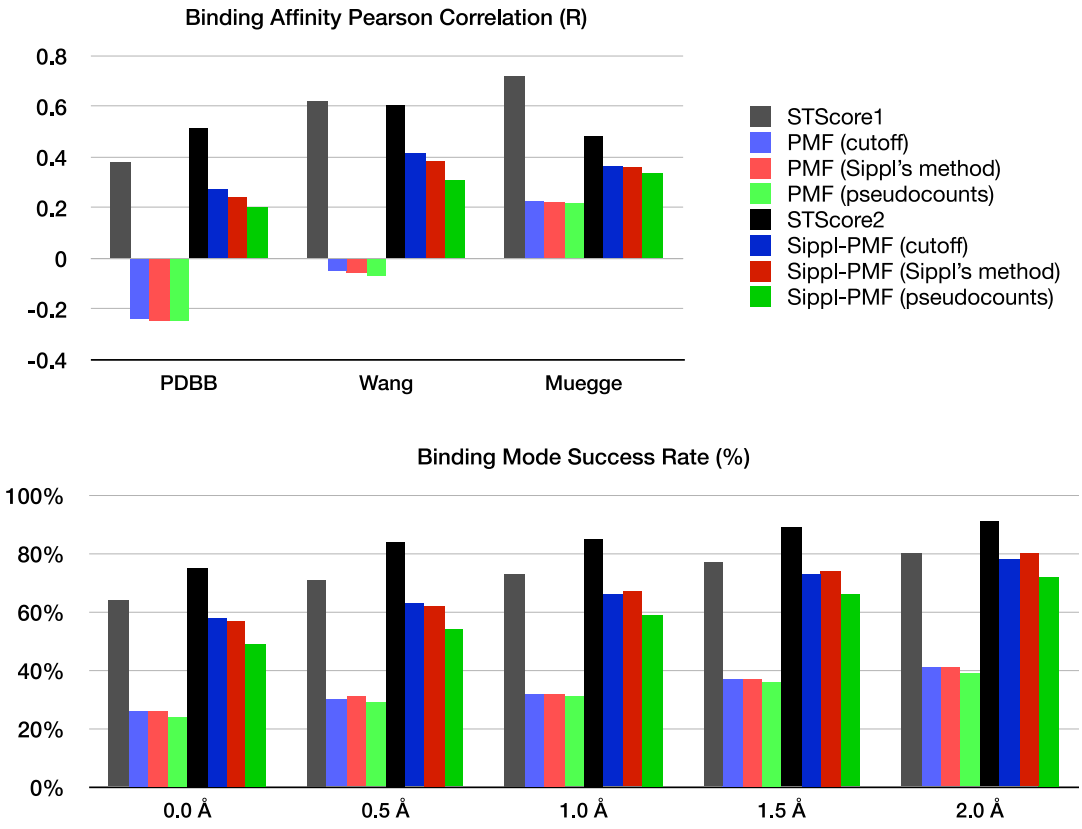

Figure 3.6: The performance of STScore1 (dark grey) and STScore2 (black), compared to other sparse data methods: the cutoff method (blue), Sippl's method (red), and the method of pseudocounts (green). STScore1 uses a simple potential of mean force (PMF) defined in equation 3.1. The performance of this PMF using the three alternative sparse data methods, along with STScore1, are shown in the first four series and colored more lightly. STScore2 uses a different potential of mean force that is based on Sippl's work (Sippl-PMF) [101]. The performance of STScore2 along with that of Sippl-PMF using the three sparse data methods are shown in the last four series and colored more darkly. The top panel gives the binding affinity prediction accuracy of the eight scoring functions using each of three testing sets: PDBbind, Wang, and Muegge. Values shown are Pearson correlation coefficients between the scores given by each scoring function and the experimentally determined binding affinities in the testing set. The bottom panel gives the binding mode prediction accuracy of the scoring functions as percent success rates, using Wang's testing set. A binding mode prediction is considered successful if the root-mean squared standard deviation between the highest-scoring pose and the native pose is within the given cutoff (0.0, 0.5, 1.0, 1.5, or 2.0 Å).

three testing sets (PDBbind, Wang, and Muegge) [6, 95, 160, 185]. The Sippl-based PMF (see Methods 3.2.6) performed much better than the PMF defined in equations 3.1 and 3.19. For the Sippl-based PMF applied to PDBbind, which is the largest testing set of the three, the binding affinity correlations were 0.241 when using Sippl's sparse data method, 0.203 when using the pseudocount method, and 0.273 when using the cutoff method. As presented above, the binding affinity correlation of PMF Score was -0.239 when using the default cutoff method for sparse data. When instead using the pseudocounts method the correlation was -0.248 and for Sippl's method the correlation was -0.245. It is worth emphasizing that the alternative sparse data methods are inheriting the bad binding affinity performance of PMF Score. STScore1 and STScore2, which benefit from a reasonable backup pair potential for regions of sparse data, gave better binding affinity predictions. The Pearson correlation was 0.381 between STScore1 and the experimentally-determined affinities of PDBbind, and for STScore2 the correlation was 0.514.

The Sippl-based PMF also gave much better binding mode predictions than PMF Score as shown in the bottom panel of Figure 3.6. Using the 2.0 Å RMSD cutoff, the Sippl-based PMF had a success rate of 78% when the cutoff method was used, 80% when Sippl's method was used, and 72% when using the method of pseudocounts. For PMF Score, the success rate was 41% when using either the cutoff method or Sippl's method, and was 39% when using the pseudocounts method.

In summary, the three sparse data methods: the cutoff method, pseudocounts, and Sippl's sparse data method, gave similar results when using either PMF Score or the PMF based on Sippl's work [101]. The Sippl-based PMF consistently performed better than PMF Score in both binding mode predictions and binding affinity predictions. In general, STScore1 and STScore2 performed better than PMF Score and Sippl-PMF Score, respectively, using any of the other three sparse data methods. STScore benefited from the influence of the FFP,

which serves as an alternative atomic pair potential for atom pair types and distances lacking training data. See Methods 3.2.2-3.2.5 for the derivation of STScore and Methods 3.2.6 for the implementations of the other sparse data methods.

3.4 Conclusions

The sparse data problem is a long-standing issue that affects all statistical potentials. In this work, we proposed a method of combining two types of scoring functions using weights based on the estimated errors in each scoring function. This combination allows the method to automatically prioritize the alternative force-field-based potential, whenever the derived potential of mean force is untrustable. This method accommodates the sparse data problem while avoiding the removal of any training data. We show that this composite potential, STScore, makes better binding mode predictions than both of the two alternative potentials, and demonstrates a new alternative to the existing sparse data methods.

One possible route for improving STScore would be to consider other, more rigorous, PMFs for protein-ligand interactions. In this work, we found that the protein-ligand PMF adopted from Sippl's work gave much better results than PMF Score, and likewise STScore2, which uses the Sippl-based PMF, gave better results than STScore1, which uses PMF Score. This motivates us to consider adapting the STScore sparse data method for use with the other knowledge-based scoring functions, to determine whether further improvements are possible.

Similarly, future work includes the adoption of other force field-based scoring functions to replace the simple force field-based scoring function used in the present study as a proof-of-concept. The sparse data method presented here can be readily generalized to other knowledge-based scoring functions, including those designed for protein structure prediction and protein-protein docking.

CHAPTER 4

Evaluating Docking Methodologies

The work in this chapter was published in Journal of Chemical Information and Modeling.[*]

Abstract

In this study, we use the recently released 2012 Community Structure-Activity Resource (CSAR) Dataset to evaluate two knowledge-based scoring functions, ITScore and STScore, and a simple force-field-based potential (VDWScore). The CSAR Dataset contains 757 compounds, most with known affinities, and 57 crystal structures. We use a scripting approach for docking preparation, and use the full CSAR Dataset to evaluate the performances of the three scoring functions on binding affinity prediction and active/inactive compound discrimination. The CSAR subset that includes crystal structures is used as well, to evaluate the performances of the scoring functions on binding mode and affinity predictions. Within this structure subset, we investigate the importance of accurate ligand and protein conformational sampling and find that the binding affinity predictions are less sensitive to non-native ligand and protein conformations than the binding mode predictions. We also find the full CSAR Dataset to be more challenging in making binding mode predictions than the subset with structures. The script files used for preparing the CSAR Dataset for docking, including scripts for canonicalization of the ligand atoms, are offered freely to the academic community.

[*]S. Z. Grinter, C. Yan, S.-Y. Huang, L. Jiang, and X. Zou (2014) Automated large-scale file preparation, docking, and scoring: evaluation of ITScore and STScore using the 2012 Community Structure-Activity Resource benchmark. *J. Comp. Chem.* 53(8): 1905–1914.

4.1 Introduction

The prospect of reliably predicting protein-ligand interactions has important implications for studies of protein function at the molecular level and for the design of therapeutic interventions [27, 29, 54, 187]. Abundant, publicly-accessible databases containing accurate protein-ligand structures and binding affinity information are invaluable tools to assess and improve the docking and scoring methods used to predict protein-ligand interactions [162, 188]. The 2012 Community Structure-Activity Resource (CSAR) Dataset, publicly released on July 3rd, contains 757 compounds with provided SMILES strings, 508 with known affinity values, 185 compounds designated to be inactive, and 57 compounds with an available protein-ligand crystal structure. These compounds span six protein targets. The compound affinities are given as either K_d, K_i, or IC_{50} and come from several different assays. For compounds with an available protein-ligand crystal structure, the CSAR Dataset provides the complex, the separated ligand and protein in their native bound conformations, and a set of unbound ligand conformations generated as MOL2 files. In this work, we use the 2012 CSAR Dataset to evaluate two knowledge-based scoring functions recently developed by our laboratory, ITScore and STScore. We also evaluate a Lennard-Jones potential as a point of reference.

ITScore [45, 46] was developed by using a statistical mechanics-based iterative method to deal with the challenging reference state problem [3, 5] faced by knowledge-based scoring functions [95, 103, 105, 174]. The approach starts by computing the native distance distributions for each atom pair type observed in the native ligand binding poses. This distribution is then compared to the same distribution but generated using a set of many decoy poses. Within the distribution that uses decoy ligand poses, each decoy is given a Boltzmann weight based on the free energy predicted for that decoy. The pair potentials can then be adjusted in a way that decreases the predicted free energy of the ligand in its native pose relative to the nearby decoy poses. This adjustment is done iteratively until the pair potentials are able

to reliably distinguish the native ligand poses from the decoys.

STScore is a knowledge-based scoring function that we developed to introduce a new method of handling the sparse data problem (not yet published). The overall approach is to combine a potential of mean force (PMF) with a simple force-field-based potential where the relative weight given to either alternative is a function of their estimated inaccuracies. We use a Bayesian statistical model to estimate the inaccuracies in the PMF due to sparse count data, allowing the method to naturally increase the influence of the force-field-based alternative for any pairs or distances lacking training data. Both the PMF and force-field-based potentials used were simple, but the overall point of STScore is to demonstrate the concept of this sparse data method, and to show that the method effectively combines the two component potentials, giving better predictions than either the PMF or force-field-based potential alone. The details of STScore will be described in a separate manuscript in preparation.

In this work, we use the 2012 CSAR Dataset to evaluate ITScore [45] and STScore. We also perform the same evaluations on a simple Lennard-Jones 6-12 potential [32, 34] as a reference. This potential, labeled VDWScore, uses the same van der Waals radii and well depths assigned for the van der Waals force in the initial potential of ITScore [45]. Consequently, ITScore and STScore are similar to VDWScore for close atom pair distances (e.g. clashes), and STScore is also similar to VDWScore for rare atom pair types.

In the following section, we will discuss the details of our evaluations of these three scoring functions. We will explain in detail how we prepared the 2012 CSAR Dataset for all the docking calculations and introduce, in Appendix A, some scripts and data files for use by future users of the CSAR Dataset. The scripts may also be of interest to others dealing with large-scale file preparation for docking. Though similar scripts are frequently used by scientists in pharmaceutical industry for large-scale docking, the scripts are not available to many academic users. In the Results, we will give special emphasis to the importance of the

native ligand and protein conformations, and discuss the challenge of handling protein and ligand flexibility in docking.

4.2 Methods

We used the 2012 CSAR Dataset to evaluate two scoring functions, ITScore and STScore. We also performed the same evaluations on a simple force-field-based potential based on the Lennard-Jones 6-12 potential. These three scoring functions have been described in the Introduction. All docking calculations were performed using the MDock 1.2 software package (http://zoulab.dalton.missouri.edu/software.htm) [45, 46]. MDock uses a rigid sampling method that is based on DOCK 4 [34]. First, spheres tangent to the protein surface are generated along the binding site. These are used to generate the initial putative ligand orientations by matching ligand atoms to the sphere points. These initial orientations are then minimized and assessed by the chosen scoring function. ITScore is the default scoring function used in MDock. For the STScore and VDWScore evaluations, the source code of MDock was modified only to include the STScore and VDWScore pair potentials and otherwise left in its original form.

4.2.1 Summary of Evaluations Performed

We evaluated the performance of ITScore, STScore, and VDWScore based on binding mode predictions and binding affinity predictions. For binding mode predictions, we computed the heavy-atom root-mean-square deviation (RMSD) between each docked ligand and the native, bound-state ligand. These predictions were then evaluated in terms of the percent success rate (where a prediction is considered successful if the top ranked ligand pose has an RMSD less than 2.0 Å). For binding affinity predictions, we computed the Pearson correlation between the docking scores and the known binding affinities of the compounds.

These correlations were computed for three separate groups according to the affinity measure provided for the compound: K_d, K_i, or IC_{50}. We also analyzed the ability of the scoring functions to distinguish between known actives and known inactives and present the data as a set of ROC curves.

We analyzed the performance of the three scoring functions using both the native structures, where available, and various structure ensembles. For the ligands, the structure ensembles were generated using the software Omega (OpenEye Scientific Software http://www.eyesopen.com) [112, 113]. For the proteins, the structure ensembles consist of either the available structures for a given protein group (for calculation on the full set of compounds in CSAR) or all of the these structures except the one bound to the ligand being docked (for subset of the CSAR Dataset that includes native protein-ligand structures). Because the full set of compounds includes many compounds with no available structure, we evaluated the scoring functions on the full set using only the Omega-generated ligand ensembles docked to the protein ensembles. For this case we evaluated the binding affinity predictions and the active/inactive compound discrimination (as ROC curves).

For the subset of the CSAR Dataset with structures, we further generated data for six other cases. Specifically, to each native protein structure we docked 1) the corresponding native ligand conformation, 2) an ensemble of conformations generated by Omega from the connection table of the ligand MOL2 file, and 3) an ensemble consisting of the one native conformation along with the set of Omega-generated conformations. For each protein ensemble, we then docked 4) the ligand in its native, bound conformation, 5) the ensemble of conformations generated by Omega from the MOL2 connection table, and 6) an ensemble of conformations generated by Omega from the SMILES string provided in the CSAR Dataset. For these six cases we evaluated the performance of the scoring functions on binding affinity predictions and binding mode predictions. These data series are depicted in detail in the

Results, and the preparation thereof is described in detail below.

4.2.2 CSAR Dataset Preparation

The CSAR Dataset contains SMILES strings for 757 compounds, 508 of which have known binding affinities. It also contains 57 protein-ligand crystal structures, some with known affinities and some without. In order to efficiently handle the data, we prepared a combined CSV datafile which contains a consistent identifier for each compound and as well as all of its associated non-structural information. This associated information includes the known binding affinity, and a label specifying whether the affinity is given as K_d, K_i, or IC_{50}. We also included the SMILES string, the type of assay used to measure the affinity, and a three-letter label specifying which protein the compound is associated with: 'CDC' for Cyclin-dependent kinase 2 bound to Cyclin A, 'CDK' for Cyclin-dependent kinase 2, 'CHK' for Checkpoint Kinase 1, 'ERK' for Extracellular signal-regulated kinase 2, 'LPX' for LpxC (a Zinc-dependent bacterial deacetylase), and 'URO' for Urokinase plasminogen activator (a serine protease).

We also included several binary labels for each compound to aid in defining useful subsets. For example, one of the labels specifies if a crystal structure is available containing the compound bound to its associated protein, and other labels specify if the compound is designated to be active or inactive. This labeling strategy allows one to easily apply commands to subsets of the CSAR Dataset defined according to these labels. The CSV datafile, which will be available at the CSAR website (http://www.csardock.org), will be of use to future users of the CSAR Dataset. More details about this datafile can be found in Appendix A.

We began with the set of protein-ligand complexes provided as MOL2 files in the CSAR Dataset and the datafile mentioned above. All other files were generated using a combination of shell scripting, Python scripting of Chimera [189] in its command-line format, and the tools distributed with the MDock software package. All statistics were done using R (http://cran.r-project.org).

Within each of the six protein groups, we aligned all of the crystal structures provided in the CSAR Dataset using Chimera's default MatchMaker settings. We also used Chimera to save the protein and the ligand into separate MOL2 files, and to convert these MOL2 files into PDB format. All the ligand files were visually inspected. The CSAR Dataset includes separated ligand and protein files, however we did not use these files because they were generated from the unaligned structures, which are less convenient for evaluating binding mode predictions.

We used Omega 2.4.3 (OpenEye Scientific Software http://www.eyesopen.com) to generate sets of ligand conformations. For each compound, the SMILES string [190] was used to generate one set of ligand conformations. For each compound with an available crystal structure (57 compounds), the MOL2 connection table was used to generate another set of conformations. The conformations generated by these two methods are the same, except in 11 cases where the SMILES strings contained ambiguous stereochemistry. In these cases, Omega handled this ambiguity by generating extra conformations to explore the stereochemistry. The conformations generated from the MOL2 file connection tables use the native stereochemistry. We used both of these methods because the former method is necessary in order to generate conformations for the full CSAR Dataset (including both compounds with and without structures) and the latter method is necessary in order to generate conformations having atom ID numbers consistent with the native structure. This ID number consistency makes the binding-mode RMSD calculations easier. Evaluations using both of these types of Omega-generated ligand ensembles give nearly identical results (as shown in the Results section).

When running Omega, we used the *-fromCT* option to ensure that the native coordinates were not being used to generate new conformations and the *-flipper true* option so that SMILES strings with ambiguous stereochemistry would be handled. We set *-strictfrags* to

true for accuracy and *-maxconfs* to 100 to keep the computational time reasonable for later docking calculations. (For SMILES strings with ambiguous stereochemistry, this conformation limit applies to each generated isomer). We also used Omega with the *-includeInput* option to provide a MOL2 file of the native conformation but with the atom IDs modified to be consistent with the other generated conformations. In summary, for each compound, Omega was used to produce an ensemble of ligand conformations and a renumbered version of the one native conformation. We also concatenated these conformations to generate an ensemble consisting of the one native conformation along with the set of Omega-generated conformations for docking.

4.2.3 MDock Docking Preparation

We used UCSF dms [189] to generate the molecular surface of each protein structure, with the default probe radius of 1.4 Å. We used Sphgen_cpp [34] to generate spheres around the whole surface of each protein. Finally, we used get_sph [34, 47], included with the MDock software package (http://zoulab.dalton.missouri.edu/software.htm), to choose spheres in the vicinity of the protein binding site, which may be defined by a PDB file. We defined this binding site broadly by concatenated all the ligand PDB files from the native aligned structures into one file for each protein group, in order to reduce bias. For native-protein docking, all binding spheres within 3.0 Å of any ligand were included.

For non-native protein docking, we used the protein ensemble method [47, 48], consistent with our approach in the CSAR benchmark exercise. To test this method on the full set, we first produced six protein ensembles by combining all of the included protein structures for each protein group (CDC, CDK, CHK, ERK, LPX, and URO). For example, the URO protein ensemble consists of the seven URO protein structures available within the CSAR Dataset: URO_4, URO_6, URO_7, URO_8, URO_9, URO_15, and URO_18. To prepare the binding sphere files for these six ensembles, we first used get_sph to generate sphere files specific to

each bound structure. For each structure, the corresponding binding sphere file contains only the spheres generated for the protein surface of that structure, restricted to the spheres within 3.0 Å of the native, bound ligand position. Then for each protein ensemble, we concatenated all of the sphere files corresponding to the protein structures that constitute the ensemble. For example, the sphere file for the URO ensemble was made by concatenating the sphere files specific to the seven URO protein structures: URO_4, URO_6, URO_7, URO_8, URO_9, URO_15, and URO_18. Finally, we used clu_sph [34, 47], included with MDock, to remove redundant spheres.

For the structure subset, we did not wish the docking calculations for any compound to benefit from the inclusion of the native protein structure for that compound within the ensemble. We therefore produced a different protein cross ensemble for each of the 56 compounds in the structure subset. Each cross ensemble includes all of the CSAR protein structures in the same protein group, except excluding the one protein structure that was bound to the compound in question. For example, the URO_6 cross ensemble consists of the six proteins structures that were not bound to the URO_6 ligand, that is: URO_4, URO_7, URO_8, URO_9, URO_15, and URO_18. This procedure was not possible for CDC (CDK2 bound to Cyclin A), for which only one crystal structure is available in the CSAR Dataset, CDC_260. Therefore, it was necessary to exclude CDC_260 from the protein cross-ensemble docking calculations performed on the structure subset, and for consistency, CDC_260 was also excluded from the native protein calculations on the structure subset. To generate a binding sphere file for each protein cross ensemble, we concatenated all of the sphere files corresponding to the protein structures that constitute that cross ensemble. For example, the sphere file for the URO_6 cross ensemble was made by concatenating the sphere files specific to the six protein structures in the URO_6 cross ensemble: URO_4, URO_7, URO_8, URO_9, URO_15, and URO_18. As with the full set ensembles, we used clu_sph [34, 47] to

remove redundant spheres.

4.2.4 Scoring Method Evaluation

For binding mode predictions, we computed the RMSD between all non-hydrogen atoms in each docked ligand and all non-hydrogen atoms in the native, bound ligand. We evaluated the binding mode predictions in terms of the percent success rate. A prediction was considered successful if the RMSD between the docked ligand and the native ligand was less than 2 Å.

To evaluate the binding affinity predictions of the scoring functions, we computed the Pearson correlation between the docked scores and the experimentally-determined binding affinities provided in the CSAR Dataset. Computing score-affinity correlations for all proteins in one group would not be appropriate, because the binding affinities for each group were obtained from assays that use different affinity measures. We therefore computed the correlation separately for each of the three affinity-measure groups: K_i, K_d, and IC_{50}. The three correlation values were then averaged to provide a generalization of the scoring functions' performance across all the proteins in the CSAR Dataset. The average was computed by taking a weighted mean of the three groups, where the weight given to a group is equal to the number of proteins in that group: two protein targets in the K_d group (considering CDC and CDK as one protein), two in the K_i group, and one protein in the IC_{50} group). We did not weight each affinity group equally, or according to the number of compounds in the group, because either of these methods would give about 30% of the weight to the IC_{50} group, which would give its single protein member, CHK, excessive influence on the averaged results. We present this weighted mean as the affinity performance measure used in the Results section. We also computed the score-affinity correlations separately for each protein group, and these correlations are provided in the Supporting Information.

In order to evaluate whether the scoring functions are able to discriminate active/inactive

compounds, we labeled the compounds as active or inactive and analyzed the receiver operating characteristic (ROC) using the ROCR package [191], available in R (http://cran.r-project.org). ROC curves, giving the ratio of the true positive rate to the false positive rate for various choices of score cutoffs, are given in the Results.

Details on the labeling of all the compounds as active or inactive are provided in Appendix A, but are briefly summarized here: Compounds were considered active if their experimentally-determined K_d, K_i, or IC_{50} value was less than $10\mu M$; compounds were considered inactive if their K_d, K_i, or IC_{50} value was known to be greater than $100\mu M$. A gap from $10-100\mu M$, within which compounds were considered neither active nor inactive, was used to reduce the risk that a compound might be labeled differently depending on the affinity measure and assay. This resulted in 491 compounds labeled active and 185 compounds labeled inactive. We provided the ROC curve for all protein groups combined, and for CDK2-Cyclin A and CDK2 separately, each of which have 22 actives and 84 inactives. Curves were not provided for the other four protein groups (CDK2, ERK2, LPXC and Urokinase) because they all had 12 or fewer labeled inactives, and two of them had no inactives.

4.3 Results and Discussion

We evaluated the performance of ITScore, STScore, and VDWScore on binding affinity predictions, binding mode predictions, and active/inactive compound discrimination. The binding affinity predictions were evaluated by computing the Pearson correlation between the docking scores and the known binding affinities for three groups, separated by affinity measure: K_d, K_i, or IC_{50}. Binding mode predictions were evaluated based on the heavy-atom root-mean-square deviation (RMSD) between each docked ligand and the native, bound ligand. Finally, ROC curves were plotted to show the active/inactive compound discrimination of the three scoring functions. These three sets of results are described as follows.

4.3.1 Binding Affinity Predictions

All binding affinity results are presented in the top panel of Figure 4.1. The y-axis values are the weighted averages of the three Pearson correlations for each affinity group (K_d, K_i, IC_{50}) where the weight given to a group is proportional to its number of protein members, as described in the Methods.

We divided these results into three different sets in order to more clearly show three comparative relationships. The first set focuses on the effect of the native versus non-native protein conformation on the affinity prediction accuracy. The second set shows the importance of the native ligand conformation versus the Omega-generated ligand conformations. The third set presents the results for both the non-native protein and Omega-generated ligand ensembles used together, using either the structure subset (i.e., the subset in which the crystal structure is provided for each protein-ligand complex) or the full set with affinities. Each set begins with the native-ligand, native-protein results in order to make the visual comparisons easier.

In the first comparison set (top-left of Figure 4.1, "native protein vs. cross ensemble"), the native-ligand, native-protein affinity results are given in one group, followed by the native-ligand, protein cross ensemble results in a second group. ITScore and STScore are found to be sensitive to the non-native protein conformation. A substantial decrease in prediction accuracy is seen when the docking is done on the protein cross ensembles instead of the native protein structures. For ITScore, the average binding affinity correlation falls from 0.81 to 0.65, and for STScore it falls from 0.61 to 0.49. No significant difference is found for the accuracy of the affinity prediction by VDWScore as a result of the protein conformation.

In the second comparison set (top-middle of Figure 4.1, "native ligand vs. Omega-generated ensemble"), the native-ligand, native-protein affinity results are given as the first group, followed by two groups that use Omega-generated ligand ensembles. The last group

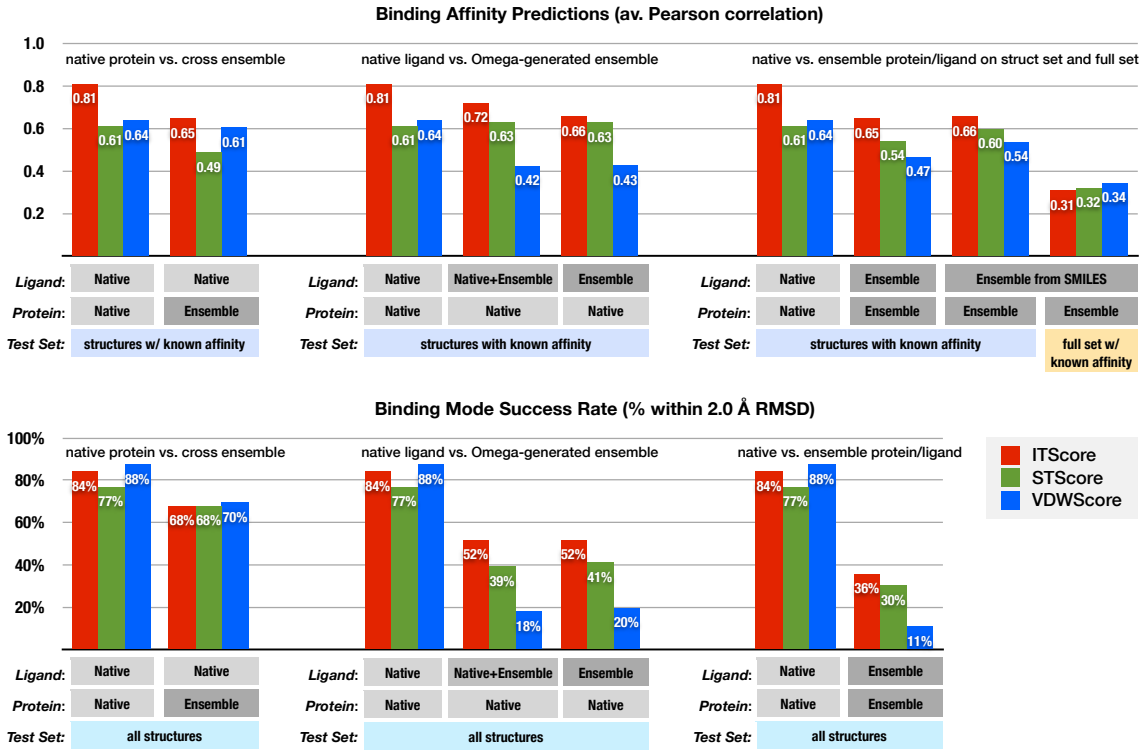

Figure 4.1: Binding affinity and binding mode predictions of the three scoring functions evaluated in this study: ITScore (red), STScore (green), and VDWScore (blue). The results are divided according to three attributes: the protein conformation (native, or ensemble), the ligand conformation (native, native+ensemble, or ensemble), and the test set used (all structures, structures with affinities, or full set with affinities). The top panel gives the binding affinity accuracy. Each y-axis value is the mean of three Pearson correlations computed for each affinity group (K_d, K_i, or IC_{50}). The affinity groups are weighted according to the number of different proteins in the group: 2.0 for K_d, 2.0 for K_i, and 1.0 for IC_{50}. The bottom panel gives the binding mode prediction results in terms of percent success rate. The binding mode is considered successfully identified by a scoring function if the lowest-scored binding mode according to that scoring function is within 2.0 Å RMSD of the native binding mode.

uses an ensemble of up to 100 Omega-generated ligand conformations ("Ligand: Ensemble"), while the middle group uses the same ensemble plus the native conformation ("Ligand: Native+Ensemble"). STScore performs similarly on all three groups with an average binding affinity correlation of 0.63 for both of the ensemble-ligand cases. ITScore shows a trend towards decreasing performance from the "Ligand: Native" group ($R = 0.81$) to the "Ligand: Native+Ensemble" group ($R = 0.72$) to the "Ligand: Ensemble" group ($R = 0.66$). VDWScore performs substantially worse on the "Ligand: Native+Ensemble" and "Ligand: Ensemble" groups with average binding affinity correlations of 0.42 and 0.43, respectively. Overall, the "Ligand: Native+Ensemble" and "Ligand: Ensemble" groups differ only slightly, suggesting that the decrease in the performance of ITScore or VDWScore from the "Ligand: Native" group is not purely the result of inadequate sampling of ligand conformations in the Omega-generated ensembles. The binding mode results (presented in the next subsection) between these two groups were also nearly identical.

On the top-right of Figure 4.1 ("native vs. ensemble protein/ligand on struct set and full set") this third comparison set shows the native-native results again, followed by three groups of protein/ligand ensemble results. The first two sets of protein/ligand ensemble results use the structure subset ("Test Set: structures with known affinity"), the same subset used for all the results presented in the first and second comparison groups. The last group uses the full set of compounds with known binding affinities. The first two groups of protein/ligand ensemble results differ only in how the ligands were generated: in the first group ("Ligand: Ensemble"), the ligand ensembles were generated by Omega from the connection table of the native ligand MOL2 file, while in the second group ("Ensemble from SMILES"), the ligand ensembles were generated by Omega from the SMILES string provided by CSAR. These two sets are the same, except for 11 compounds with stereochemical ambiguities in the SMILES strings; for these compounds, extra conformations were generated by Omega to explore

the stereochemical space. The binding affinity results between these two ligand generation methods were close, with ITScore having average binding correlations of 0.65 and 0.66 on the two groups respectively. For STScore, these values were 0.54 and 0.60, and for VDWScore they were 0.47 and 0.54. We conclude that the large difference in performance between the first ensemble-ensemble group on the structure subset and the last ensemble-ensemble group on the full set was due to the difficulty of the full test set versus the structure subset, and not due to the method of generating the ensembles of ligand conformations.

In the first two groups of ensemble-ensemble results, a decrease in performance is seen for ITScore relative to the native-native results. This decrease, from $R = 0.81$ to $R = 0.65$ is the same decrease seen for the native-ligand, ensemble-protein case ($R = 0.65$), and the ensemble-ligand, native-protein case ($R = 0.66$). A similar decrease is seen for VDWScore (from $R = 0.61$ to $R = 0.54$) although the decrease is insignificant for the "Ensemble from SMILES" case ($R = 0.60$). Overall, the performance of STScore is close between the native-native ($R = 0.61$) and ensemble-ensemble cases ($R = 0.54$ for ligand conformations from the connection table, and $R = 0.60$ for the conformations from the SMILES strings).

In the last group of protein/ligand ensemble results, the full set of compounds (as ensembles generated from the CSAR-provided SMILES strings) were docked to the non-native protein ensembles for each protein group. As with the other groups, the y-axis value is the weighted average of the Pearson correlations for the K_d, K_i, and IC_{50} subsets, where the weight given to each group is proportional to the number of proteins within the group (two for K_d, two for K_i, and one for IC_{50}). A decrease in performance is seen for all three scoring functions on the full set of ensemble-ensemble results compared to the structure subset. For ITScore the average binding affinity correlation was 0.31, for STScore it was 0.32 and for VDWScore it was 0.34. For ITScore, which performed better than the other two scoring functions in every other case, this decrease in the average binding affinity correlation was

large (from $R = 0.65$ for the structure subset to $R = 0.31$ for the full set with affinities). As discussed above, the ligand ensembles generated from the MOL2 connection table ("Ligand: Ensemble") and the SMILES ligand ensembles ("Ligand: Ensemble from SMILES") give close results. This consistency in results suggests that the substantial decrease in performance seen with the full CSAR Dataset is due to greater difficulty in the full set itself. The full CSAR Dataset includes many compounds with similar activity values, and these activity values are less consistently related to the basic ligand parameters than is the case for the structure subset. For example, the affinity values in the structure subset are moderately correlated to the number of atoms in each ligand; however, for the full set, this correlation is very weak.

Overall, ITScore made better binding affinity predictions than the other two scoring functions in every case except the ensemble-ensemble evaluation on the full set of compounds with affinities. The binding affinity prediction accuracy of STScore and of VDWScore were similar, with one doing better than the other in about half of the cases. Given the crudely simple functional form of VDWScore, its comparable affinity performance was unexpected. We suspect that its results may have been improved by the binding pocket hydrophobicity of several of the proteins in the 2012 CSAR Dataset, as depicted in Figure 4.2. While the surrounding regions can be seen to contain many hydrophilic residues, the actual contact residues are predominately hydrophobic for four of the proteins, and mixed for the other two proteins.

4.3.2 Binding Mode Predictions

All binding mode results are presented in the bottom panel of Figure 4.1. The y-axis values are the percent of compounds for which the top ranked binding mode and the native binding mode are within 2.0 Å RMSD (root-mean-square deviation), excluding hydrogen atoms. As with the affinity results, the binding mode results are divided into three sets in order to more clearly show the three comparative relationships: "native protein vs. cross ensemble,"

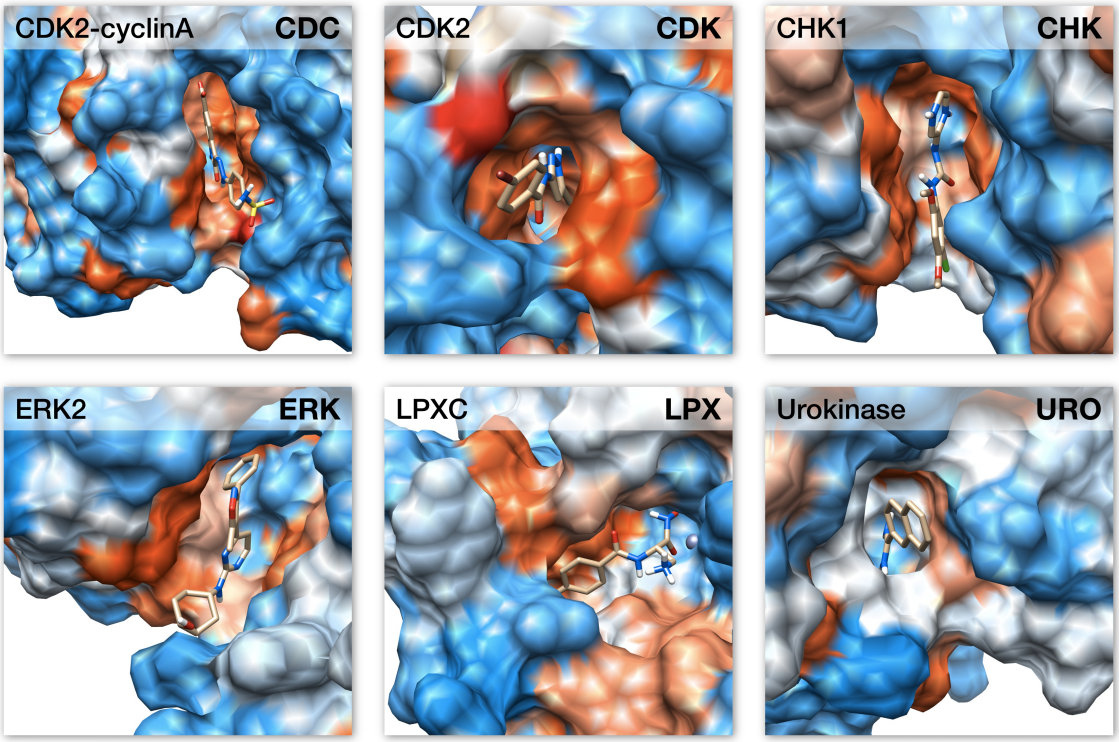

Figure 4.2: Example binding modes for each of the six protein groups. The protein surface around each residue is colored according to the Kyte and Doolittle scale of hydrophobicity [192]. Hydrophilic residues are colored blue while more hydrophobic residues are colored orange or red. The figure was generated using UCSF Chimera 1.6.2.

"native ligand vs. Omega-generated ensemble," and "native vs. ensemble protein/ligand." Because evaluating the binding mode requires knowledge of the native ligand position, the binding mode evaluations were restricted to the structure subset of compounds ("Test Set: all structures").

In the first comparison set (bottom-left of Figure 4.1: "native protein vs. cross ensemble"), the native-ligand, native-protein binding mode results are given in the first group, followed by the native-ligand, ensemble-protein results in the second group. All scoring functions show a decrease in binding mode predictions as a result of using the protein cross ensembles instead of the native protein conformations. For ITScore, the success rate drops from 84% to 68%. For STScore the native-native success rate is 77%, and drops to 68% when the protein ensemble is used instead. VDWScore does well in the native case, with a success rate of 88%, and the success rate drops to 70% with the protein ensemble.

In the second comparison set (bottom-middle of Figure 4.1: "native ligand vs. Omega-generated ensemble"), it is shown that the native ligand conformation is very important for successful binding mode predictions when testing the three scoring functions on the 2012 CSAR Dataset. All three scoring functions show a large decrease in performance when using an Omega-generated ligand ensembles ("Ligand: Native+Ensemble" or "Ligand: Ensemble") rather than the native ligands, and this decrease in performance is substantially larger than the decrease seen when using the protein cross ensembles versus the native protein conformations. For ITScore the success rate drops from 84% to 52% for both Omega-generated ligand ensembles. For STStore the success rate drops from 77% to 39% for the "Ligand: Native+Ensemble" case, and to 41% for the "Ligand: Ensemble" case. For VDWScore the decrease is the greatest, from 88% to 18% and 20%, suggesting that this scoring function (which does not consider electrostatics, explicitly or implicitly) is especially bad at distinguishing the native ligand conformation.

On the bottom-right of Figure 4.1 ("native vs. ensemble protein/ligand") is the third comparison set. It shows the native-native results as a point of comparison followed by the protein/ligand ensemble results. As with the ensemble-ligand, native-protein results, the ensemble-ensemble binding mode predictions are much worse than the native-native binding mode predictions for all three scoring functions. Between the ensemble-ligand, native-protein results and the ensemble-ligand, ensemble-protein results, a trend of decreasing performance is seen for all three scoring functions. The success rates decrease to 36% and 30% for ITScore and STScore, respectively. VDWScore, whose performance was already only 20% in the ensemble-ligand, native-protein case, does particularly poorly in the ensemble-ensemble case, with a success rate of 11%.

To summarize, when the native, bound ligand conformation is used, all three scoring functions give similar binding mode performance. For the native-native case, STScore performs slightly worse than the other two scoring functions. Whenever the Omega-generated ligand ensemble is used, the binding mode performance of all the three scoring functions decreases substantially. For VDWScore this decrease is very large, because the shape complementarity is not sufficient for docking in the absence of native ligand conformations, even with the pockets that are mostly hydrophobic.

4.3.3 Active/Inactive Compound Discrimination

The performance of the three scoring functions on active/inactive compound discrimination are presented as a set of three ROC curves for each scoring functions in Figure 4.3. These curves give the ratio of the true positive rate to the false positive rate for each possible choice of docking score cutoff. The labeled values on each curve give example cutoffs. For example, at the top-left, the ROC curve titled "ITScore on Full Set" labels the curve at cutoffs of $ITScore = -65$, $ITScore = -60$, $ITScore = -55$, $ITScore = -50$, and $ITScore = -45$.

The first set of ROC curves (left panels) give the ROC curves for ITScore, STScore, and

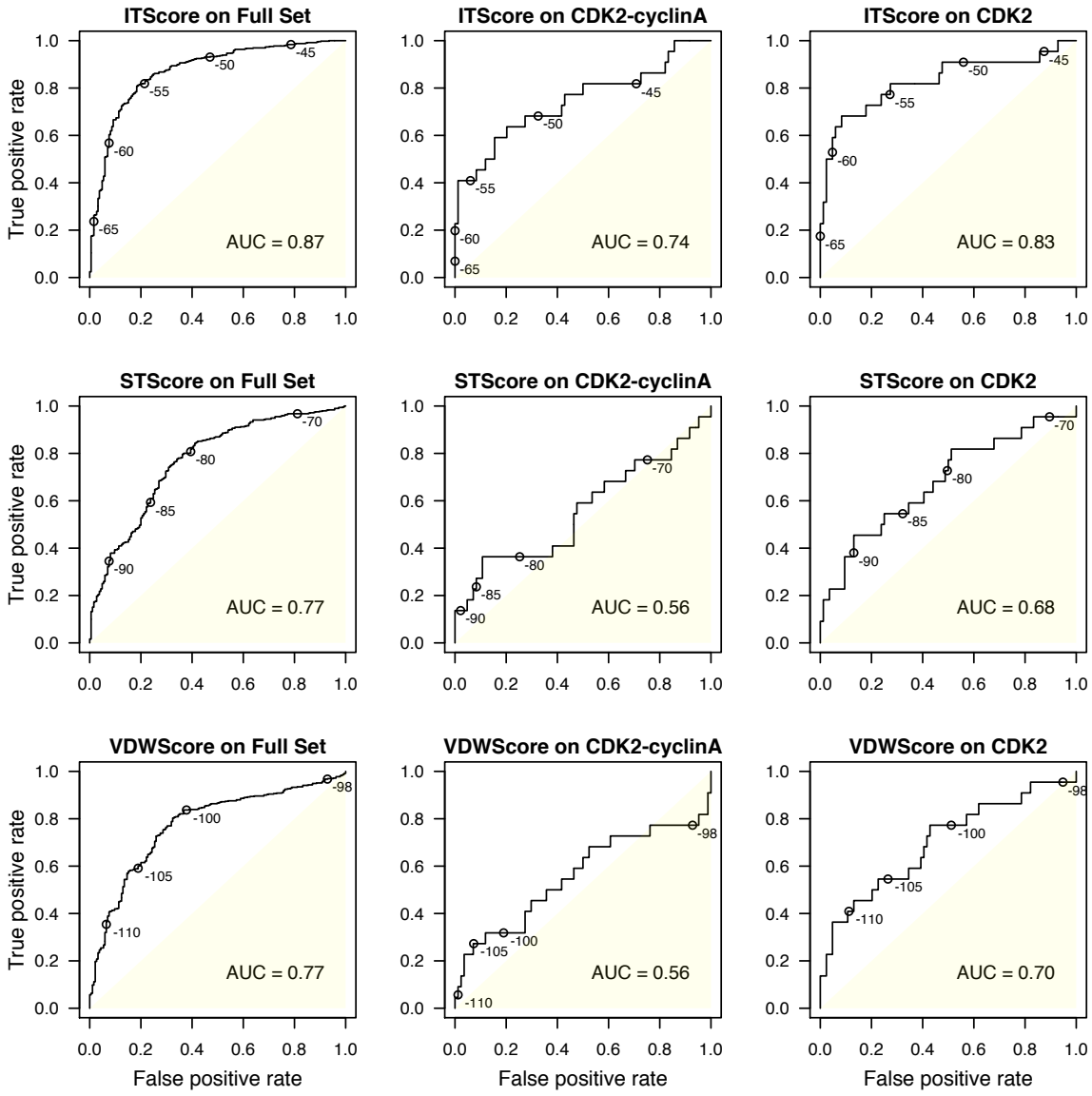

Figure 4.3: The active/inactive compound discrimination of ITScore (row 1), STScore (row 2), and VDWScore (row 3), presented as receiver operating characteristic (ROC) curves. These curves give the ratio of the true positive rate to the false positive rate for each possible choice of docking score cutoff. The labeled values on each curve give example cutoffs. The first column gives the ROC curves for the full set of active and inactive compounds. This column combines compounds from all six protein groups and tests the ability of each scoring function to discriminate between the actives and inactives of different proteins. The second and third columns give ROC curves for CDK2-Cyclin A and CDK2. Specific ROC curves for other protein groups were not provided because they all had 12 or fewer inactives.

VDWScore on the full set of active and inactive compounds. This combines compounds from all six protein groups, testing the ability of each scoring function to discriminate between the actives and inactives of different proteins. The second and third columns give ROC curves for specific protein-group subsets of the full set. Only CDK2-Cyclin A and CDK2 are given because these are the only protein groups for which more than 20 actives and 20 inactives were available. The other protein groups have 12 or fewer inactives.

ITScore performs well in all three evaluations with an area under the curve (*AUC*) of 0.87 on the full set, and 0.74 and 0.83 for CDK2-Cyclin A and CDK2, respectively. STScore and VDWScore do not perform as well as ITScore on the full set ($AUC = 0.77$ for the two). On the individual protein evaluations, STScore and VDWScore perform poorly. For CDK2-Cyclin A, the *AUC* equals 0.56 for both, narrowly better than random selection. For CDK2, the *AUC* was 0.68 for STScore and 0.70 for VDWScore.

4.3.4 Summary of Results

In summary, ITScore performed relatively well in all binding affinity predictions on the structure subset. It also performed well in active/inactive compound discrimination. VDWScore performed less well than ITScore in the binding affinity predictions on the structure subset, but slightly better than ITScore in non-native protein/ligand ensemble binding affinity predictions on the full set.

In binding mode predictions, ITScore and VDWScore performed similarly when the native ligand conformation was used. When the Omega-generated ligand ensembles were used instead, ITScore performed much better than VDWScore. For all scoring functions, the generated ligand conformation ensembles gave worse binding mode results than the native ligand conformation, and in these cases VDWScore did particularly poorly.

Overall, in the Omega-generated ligand ensemble cases, the performance of STScore was consistently better than VDWScore in both binding mode and binding affinity predictions.

The exception was the full set with known affinities: in this case the affinity predictions of STScore and ITScore were slightly worse than those of VDWScore.

In the native-native case of the CSAR structure subset, the binding affinities had average Pearson correlations of 0.81, 0.61, and 0.64 with the docking scores given by STScore, ITScore, and VDWScore, respectively. For the ensemble-ensemble case on the structure subset, these values fell to 0.65, 0.54, and 0.47. In the native-native case, the binding mode prediction success rates were 84%, 77%, and 88% for STScore, ITScore, and VDWScore. These values fell to 36%, 30%, and 11% in the ensemble-ensemble case (where the success criterion for binding mode prediction is that the top ranked binding mode and the native binding mode are within 2.0 Å RMSD). We also found the full CSAR dataset to be more challenging in making binding mode predictions than the subset with structures. For the full set of compounds with known affinity, the binding affinities had average Pearson correlations of 0.31, 0.32, and 0.34 with the docking scores given by STScore, ITScore, and VDWScore, respectively. For the active/inactive compound discrimination all the scoring functions performed better. In evaluating the ROC on the full set, the area under the curve was 0.87 for ITScore, and 0.77 for both STScore and VDWScore.

4.4 Conclusions

Our data supported some of our previous conclusions about ITScore and STScore, and in other cases contradicted our expectations. Our results are consistent with our previous conclusion [161, 193] that ITScore's iterative method of dealing with the reference state problem is able to substantially increase binding mode and binding affinity predictions compared to its initial potential (which is similar to STScore for abundant pair types and similar to VDWScore for close atom pair distances). All three scoring functions were tested using MDock, which positions an ensemble of pre-generated/rigid ligand conformations without further confor-

mational sampling.

STScore is similar to the initial potential of ITScore for abundant pair types (which account for the majority of interactions) but is also similar to VDWScore for rare atom pair types. Consequently, it fell within our expectations that the performance of STScore was often in-between that of ITScore and VDWScore. Evidently, the knowledge-based component of STScore improves its binding mode predictions compared to its force-field based component alone, which is a slightly modified version of VDWScore. Its binding mode predictions were still not as good as ITScore, although our unpublished results suggest that ITScore-like iterations can further improve the binding mode predictions of STScore.

Considering its simple functional form, it was initially surprising that the binding affinity predictions of VDWScore exceeded STScore in some cases and ITScore in one case. It is possible that its performance was enhanced by the hydrophobicity of the protein binding pockets in the 2012 CSAR Dataset. This would make the task easier for VDWScore, because it must rely primarily on the shape complementarity of the protein and the ligand. In support of this view (that the performance of VDWScore is highly dependent on shape complimentarily), there was a significantly larger difference in the binding mode performance of VDWScore depending on whether the native or Omega-generated ligand conformation ensemble was used. This difference was much greater for VDWScore than for ITScore or STScore. When docking the native ligand conformation, VDWScore gave quite accurate binding mode predictions, but in those cases that use the Omega-generated ligand ensembles, its binding mode predictions were very poor. So in summary, these results suggest that the knowledge-based aspect of ITScore and STScore is able to increase their binding mode predictions beyond that of VDWScore by implicitly including other types of interactions. The contribution of other interactions is especially important when the native conformation of the ligand is unknown. Nevertheless, STScore and ITScore leave much room for improvement in binding affinity and

binding mode predictions when using the non-native protein/ligand conformations.

In general, our scoring experiments with the full set of compounds from the 2012 CSAR benchmark suggest that it is interesting yet challenging to predict binding modes and affinities without the knowledge of native protein and ligand conformations. The van der Waals scoring function may be used as a reference for scoring comparison; van der Waals performs much better on predictions for native protein and ligand conformations than for non-native conformations. The use of the pre-generated ligand conformations seems to lower the success rates significantly more than the use of the non-native protein conformations for binding mode predictions. The corresponding difference is less for binding affinity predictions. This phenomenon may be due to the fact that the main conformational changes of the proteins are side chain flexibility in 2012 CSAR Dataset. Future work may include adapting these scoring functions and docking methods for use with on-the-fly ligand conformational sampling and side chain rotamer sampling.

Supporting Information

Additional figures and a CSV datafile (see Appendix A) are provided in the Supporting Information. This information is available free of charge via the Internet at http://pubs.acs.org.

CHAPTER 5

Application to Breast Cancer

The work in this chapter has been published in Journal of Molecular Graphics and Modelling[*] *and includes experimental data from the laboratory of co-corresponding author Dr. Salman M. Hyder generated by co-first author Dr. Yayun Liang.*

Abstract

Inverse docking is a relatively new technique that has been used to identify potential receptor targets of small molecules. Our docking software package MDock is well suited for such an application as it is both computationally efficient, yet simultaneously shows adequate results in binding affinity predictions and enrichment tests. As a validation study, we present the first stage results of an inverse-docking study which seeks to identify potential direct targets of PRIMA-1. PRIMA-1 is well known for its ability to restore mutant p53's tumor suppressor function, leading to apoptosis in several types of cancer cells. For this reason, we believe that potential direct targets of PRIMA-1 identified *in silico* should be experimentally screened for their ability to inhibit cancer cell growth. The highest-ranked human protein of our PRIMA-1 docking results is oxidosqualene cyclase (OSC), which is part of the cholesterol synthetic pathway. The results of two followup experiments which treat OSC as a possible anti-cancer target are promising. We show that both PRIMA-1 and Ro 48-8071, a known potent OSC inhibitor, significantly reduce the viability of BT-474 breast cancer cells relative to normal mammary cells. In addition, like PRIMA-1, we find that Ro 48-8071 results in increased

[*]S. Z. Grinter, Y. Liang, S.-Y. Huang, S. M. Hyder, and X. Zou (2014) An inverse docking approach for identifying new potential anti-cancer targets. *J. Mol. Graph. Mod.* 29(6): 795–799.

binding of mutant p53 to DNA in BT-474 cells (which highly express p53). For the first time, Ro 48-8071 is shown as a potent agent in killing human breast cancer cells. The potential of OSC as a new target for developing anticancer therapies is worth further investigation.

5.1 Introduction

Inverse docking, first proposed in 2001 by Chen et al. [153] refers to computationally docking a specific small molecule of interest to a library of receptor structures. The technique may be used to identify new potential biological targets of known compounds [194–196], or to identify targets for compounds among a family of related receptors [197]. The technique has shown success in distinguishing between homology models of receptors [197]. The technique may also be used to generate a compound's predicted pharmacological profile [198], or to generate a virtual selectivity profile that characterizes the promiscuity of the inhibitors [199]. Given the multi-faceted nature of a pharmacologically active compound's biological effects, inverse docking is especially helpful, because it may generate new hypotheses for the action mechanism.

Our docking software package, MDock, can be used for inverse docking, as demonstrated in the present work (http://zoulab.dalton.missouri.edu/software.htm). MDock uses a novel scoring function, ITScore, which was generated using an iterative method of deriving pair interaction potentials that avoids the problem of defining a specific reference state [45]. For the first time, the full energy landscape (both native and non-native modes) is considered in the potential derivation using a physics-based global iterative function. ITScore's binding pose and affinity predictions have been extensively evaluated using diverse test sets prepared by other labs [45, 46]. ITScore was also assessed using enrichment tests for virtual database screening against four target proteins [46]. In the present study, we test the ability of MDock on textitin silico inverse screening applications.

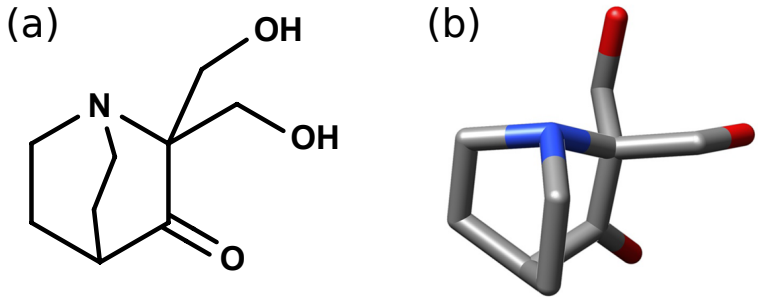

Figure 5.1: (a) Chemical structure, generated using MarvinSketch 4.1.0 (http://www.chemaxon.com), and (b) 3D structure of PRIMA-1. Hydrogen atoms are omitted from the 3D structure for clarity.

Specifically, we aim at searching for potential protein targets of PRIMA-1. Found from high-throughput screening, PRIMA-1 (p53 reactivation and induction of massive apoptosis, shown in Figure 5.1), is a small molecule capable of activating mutant p53 protein, restoring its ability to bind to DNA as well as the tumor suppressor function associated with wild-type p53 [8, 200]. This effect has been demonstrated in vitro and in vivo, and has been shown to trigger massive apoptosis in several types of human breast cancer cells [201, 202]. PRIMA-1 is also known to stimulate expression of p21 and other p53-dependent promoters in mutant p53 breast cancer cell lines. p53's importance as a potential agent against cancer is well-established. Nevertheless, while specific mechanisms have been proposed for PRIMA-1's mutant p53 reactivation effect [8, 203, 204], none have gained wide acceptance and the question remains unsettled. For this reason, we consider PRIMA-1 well suited as the subject of an inverse docking study.

In this work, we used the inverse-docking approach to screen for potential molecular targets of PRIMA-1. The objective is to guide future assays of the inhibitors of these predicted targets for their efficacy in inhibiting tumor cell proliferation, as such results may lead to potential cancer treatments, as well as provide clues regarding PRIMA-1's action mechanism. We used MDock to perform this study. In support of our approach, here we present the first stage results of our assays of Ro 48-8071, a known potent inhibitor of oxidosqualene

cyclase (OSC) [205, 206], the highest-ranked human protein of our *in silico* study. We show that Ro 48-8071 is a novel potent agent in selectively reducing the viability of BT-474 cells, a mutant-p53 human breast cancer cell line. In addition, we found that Ro 48-8071 increases p53-DNA binding in BT-474 cells, an effect which is also characteristic of PRIMA-1 [200]. BT-474 cells are known to overexpress p53, even in the absence of cytotoxic stress [207].

5.2 Methods

5.2.1 In Silico Screening

We used our protein-ligand docking software package MDock [45, 46] (http://zoulab.dalton.missouri.edu/software.htm) to dock PRIMA-1 into many potential drug targets. Although MDock is sufficiently computationally efficient for PDB-wide database screening, we chose to start with the well-characterized Potential Drug Target Database (PDTD), which at the time of use contained about 1100 experimentally-determined structures of 830 actual or suspected drug targets (http://www.dddc.ac.cn/pdtd) [155]. We also used the PDTD's binding site definitions, which in most cases are based on the set of amino acid residues that are within 6.5 Å of the bound ligand. OMEGA Version 2.2.1 was used to generate conformations of PRIMA-1 for flexible-ligand docking (OpenEye Scientific Software Inc., Santa Fe, NM http://www.eyesopen.com) with the *rms* parameter set to 0.1 Å, *maxconfs* to 1000000, *maxconfgen* to 10000000, and *ewindow* to 10 [112, 113]. As PRIMA-1 has few rotatable bonds, this only resulted in 42 generated ligand conformations. Each of these conformations was docked to each protein as a rigid body.

Our docking procedure is described in detail in previous publications [45–49] and in the tutorial of MDock. Briefly, for each protein in the database, a molecular surface of the binding site was generated, along with the associated sphere points representing potential initial positions for ligand atom centers [34, 208]. Ligand atoms were matched to these

sphere points and orientations were sampled and ranked by our knowledge-based scoring function, ITScore [45, 46]. All of MDock's default parameters were used in this work, with the exception of *write_score_total*, which was set to 1 so that only the highest-scoring orientation is recorded when each protein/PRIMA-1-conformation pair is docked as a rigid body. We then ranked each protein according to the lowest ITScore (corresponding to the highest predicted affinity) recorded for any of the 42 PRIMA-1 conformations that were docked to it. Because PDTD contains redundant experimental structures of the same protein [155], we clustered the resulting docked structures into groups sharing ≥ 90% sequence identity. We then ran a BLAST search [209] in order to map the PDTD proteins, which come from various species, to human gene sequences. Inhibitors of the top human or human analogue proteins were considered candidate anti-cancer agents for assay. A flowchart of our procedure is shown as Figure 5.2.

5.2.2 Cell Viability Assay

We used the Sulforhodamine B (SRB) assay [210, 210–212] to evaluate the effect of the OSC-inhibitor Ro 48-8071 on the viability of breast cancer cells. This cell protein dye-binding assay determines the protein content in surviving cells as an index to determine cell growth, viability, and survival [210, 210]. Briefly, BT-474, T47D, and AG11132A cells were seeded into 96-well plates and incubated overnight at 37°C with 5% CO2. The culture medium was removed after 24 h and cells were washed with DMEM/F12 medium and then treated with various concentrations of Ro 48-8071 or PRIMA-1 in 5% FBS DMEM/F12 medium for 24 hours. Surviving or adherent cells were fixed in situ by withdrawing the growth medium, adding 100 µl PBS and 100 µl 50% trichloroacetic acid and then incubating at 4°C for one hour. Cells were washed with ice-cold water, dried at room temperature (RT), and then stained with 50 µl 4% SRB for eight minutes at RT. Unbound dye was removed by washing five times with cold 1% acetic acid and plates were dried at RT. Bound stain was solubilized with 150

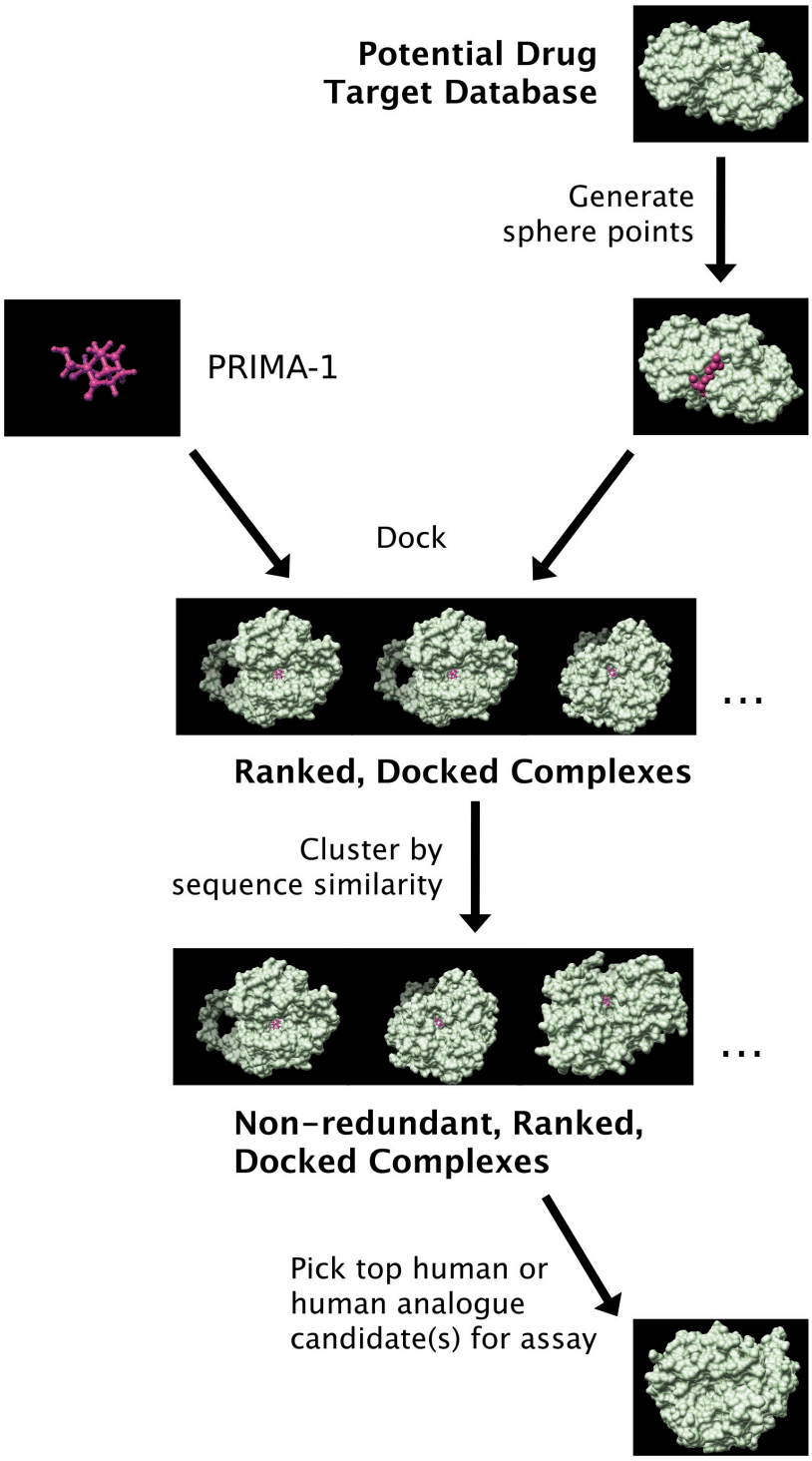

Figure 5.2: A flowchart illustrating the inverse docking and assay approach used in this work.

µl of 10 µM Tris buffer, and the absorbance of samples was read at 520nm with a SpecTRA MAX 190 microplate reader (Molecular Devices, Sunnyvale, CA). Six wells were used for each concentration and each experiment was performed twice. BT-474 and T47D breast cancer lines were obtained from ATCC (Manassas, VA), and the AG11132A normal mammary cell line was purchased from Coriell Institute for Medical Research (Camden, NJ). BT-474 and T47D cells were grown in phenol red-free DME/F12 medium (Invitrogen Corporation; Carlsbad, CA) and supplemented with 10% fetal bovine serum (FBS; Sigma-Aldrich, St. Louis, MO). AG11132A cells were grown in serum free MEBM (Mammary Epithelium Basal Medium) medium (Lonza, Walkersville, MD) with supplementary 2mM L-glutamine. PRIMA-1 was purchased from Tocris Bioscience (Ellisville, MO). Ro 48-8071, sulforhodamine B, and other chemicals were purchased from Sigma-Aldrich (St. Louis, MO). The purity of Ro 48-8071 was ≥ 98% as determined by HPLC (Sigma-Aldrich data). The purity of PRIMA-1 was 99.8% as determined by HPLC (Tocris data sheet).

5.2.3 p53 Activation Assay

In preparation for the assay, BT-474 cells were grown in DMEM/F12 medium supplemented with 5% FBS overnight. Cells were washed with PBS once and treated with 50 µM PRIMA-1 or 25 µM Ro 48-8071 for 1 hour at 37°C. p53 activation was assessed using the TransAM p53 Transcription Factor Assay kit (Active Motif, Carlsbad, CA) according to the manufacturer's protocol. A summary of the procedure follows. The kit provides 96-well plates coated with an oligonucleotide that contains the p53 consensus DNA binding site. 2.5 µg of nuclear extracts (prepared according to a nuclear extract kit provided from Active Motif) were incubated with this oligonucleotide. Bound p53 was detected by adding the anti-p53 antibody (1:1000) followed by addition of the secondary antibody (1:1000) that is conjugated to horseradish peroxidase. Absorbance was read at 450nm in a Spectra MAX 190 Microplate Reader (Molecular Device, Sunnyville, CA). MCF-7 nuclear extract treated with H2O2, provided with the

TransAM kit, was used as a positive control.

5.2.4 Statistical Analysis

Differences among groups were tested using one-way analysis of variance (ANOVA) with repeated measures over time. Values are reported as mean ± SE. When ANOVA indicated a significant effect (F-ratio, $p < 0.05$), the Student-Newman-Keuls multi-range test was used to compare the means of the individual groups. The statistics were conducted using the SigmaStat software (version 3.5).

5.3 Results and Discussion

5.3.1 In Silico Screening

After docking PRIMA-1 to each structure of the Potential Drug Target Database (PDTD), we ranked the proteins according to their predicted affinity, based on our knowledge-based scoring function, ITScore. We searched for human proteins that are analogous to the best-scoring (i.e., lowest-scoring or tightest-binding) proteins in our docking results, using a cutoff of 30% sequence identity. Among these ten best-scoring proteins, one of them is a human protein, the X-ray crystallographic structure of human OSC (PDB entry: 1W6K) [213, 214]. In Figure 5.3, OSC (green) is shown docked with PRIMA-1 (magenta) along with the potent OSC-inhibitor, Ro 48-8071 (yellow) [206, 215]. The binding pose indicated by docking PRIMA-1 into OSC partially overlaps the binding pose of Ro 48-8071 shown in the crystal structure. We also found that docking Ro 48-8071 to this pocket reproduces the native binding orientation shown in the crystal structure (RMSD = 0.25 Å). The score for PRIMA-1 calculated with ITScore was -45.5 and the score for Ro 48-8071 in its crystallographic position was -102.8. Approximately, this difference in score corresponds to a 6 kcal/mol difference in predicted binding affinity between the two compounds.

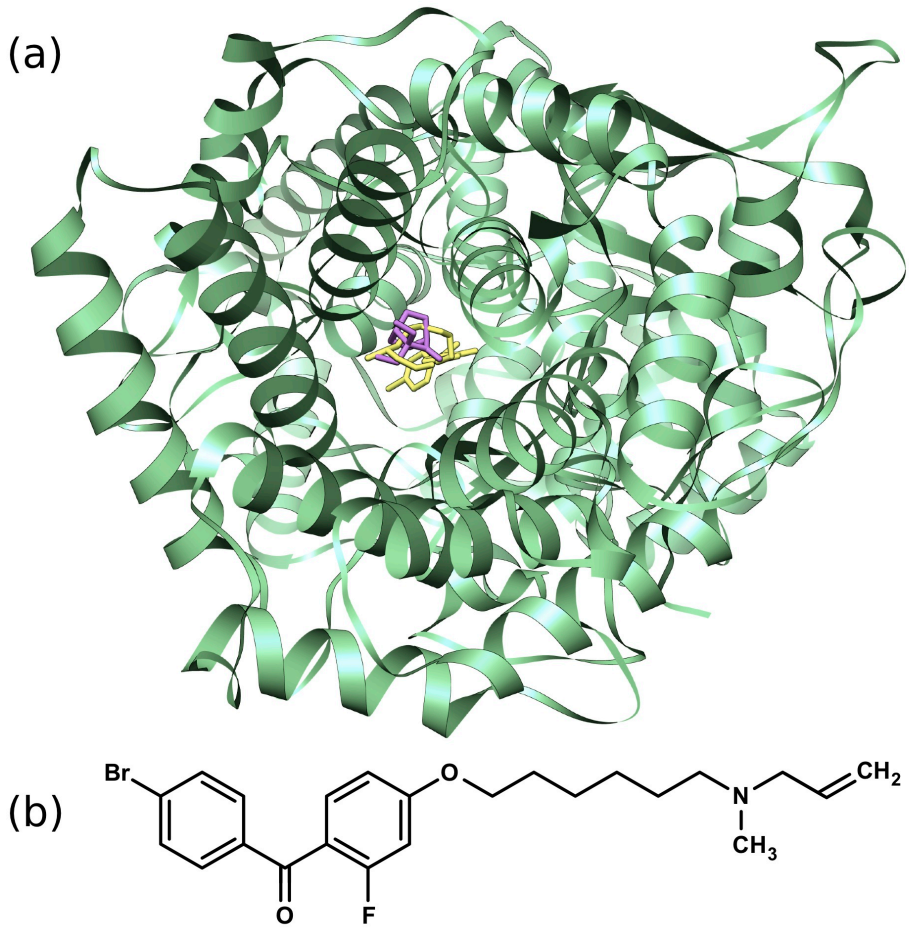

Figure 5.3: (a) Ribbon depiction of oxidosqualene cyclase (OSC), identified as a possible target of PRIMA-1, generated using Chimera 1.4.0 [189]. PRIMA-1 (magenta) is shown in its docked position along with the partially overlapping position of the OSC inhibitor Ro 48-8071 from the crystal structure (yellow). Hydrogen atoms are omitted for clarity. (b) Chemical structure of Ro 48-8071.

	PRIMA-1	RO 48-8071
Binding score	-38.7	-61.3
VDW score	-24.3	-52.2
Electrostatics score	-14.3	-9.2

Table 5.1: The binding energy scores and individual energy components of PRIMA-1 and Ro 48-8071 docked to OSC, calculated with UCSF DOCK 6.0 [36].

To further compare the similarities and differences between the interactions involved in PRIMA-1 binding and Ro 48-8071 binding, we attempted to decompose the total energy scores into different energy components. Unfortunately this cannot be done with ITScore because the potential function in ITScore derived for each atom pair combines different energetic contributions into a single distance-dependent function. We therefore used the force field scoring function [32] provided in DOCK 6.0 (UCSF, http://dock.compbio.ucsf.edu/) [189] to analyze the natures of the interactions involved in binding of PRIMA-1 and Ro 48-8071 to OSC, by calculating the contributions of different energy terms to the total binding scores.

Specifically, the force field scoring function in UCSF DOCK 6.0 is composed of two energy terms, a van der Waals (VDW) term using Lennard-Jones 6-12 potentials and a Coulombic electrostatic energy term using a distance-dependent function for the dielectric constant of water. Table 1 lists the binding energy scores of PRIMA-1 and Ro 48-8071 and the corresponding contributions of different energy components. It can be seen from the table that Ro 48-8071 (-61.3) has a lower/better binding score than PRIMA-1 (-38.7), which is consistent with the afore-mentioned results calculated with ITScore. Table 1 also suggests that the VDW interactions contribute to the binding energies significantly more than the electrostatic interactions for both PRIMA-1 and Ro 48-8071, though the contribution of the VDW interaction term is more dominant for Ro 48-8071 than PRIMA-1. The strong VDW interactions for Ro 48-8071 arise from its highly hydrophobic fragments such as aromatic rings and aliphatic chains.

Since PRIMA-1 inhibits cell growth in mutant-p53 tumor cell lines, we decided to deter-

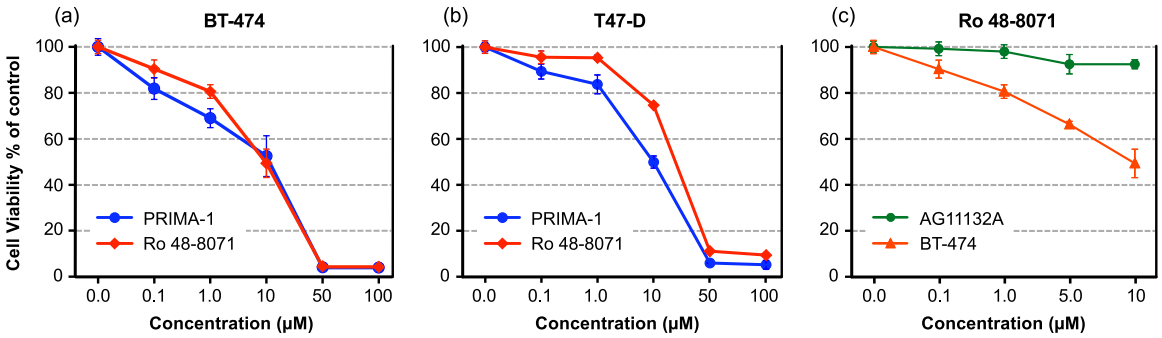

Figure 5.4: The effect of Ro 48-8071 on breast cancer and normal mammary cell viability. BT-474 (1.0x104 / well), T47-D (0.6x104 / well), and AG11132A cells (0.7x104 / well) were seeded into a 96-well plate overnight, and cells were washed and treated with the indicated concentration of Ro 48-8071 or PRIMA-1 for 24 hours. Cell growth and viability were determined by the SRB assay described in Methods. The OSC-inhibitor Ro 48-8071 and PRIMA-1 significantly inhibit the viability of BT-474 (a) and T47-D (b) cells in a dose-dependent manner, and there is significantly less inhibition of normal mammary AG11132A cell viability shown in (c). (Data provided by co-corresponding author Dr. Salman M. Hyder and co-first author Dr. Yayun Liang.)

mine whether the potent OSC-inhibitor Ro 48-8071 would have a similar anti-cancer effect.

5.3.2 Cell Viability Assay

The SRB assay showed that Ro 48-8071 dramatically destroys BT-474 human breast cancer cells, exhibiting a dose-response relationship similar to that of PRIMA-1. IC50 was approximately 10 μM for both compounds. The OSC-inhibitor also suppressed the growth of a second human breast cancer cell line, T47D. The data for both cell lines are shown in Figure 5.4 (a) and (b). Using the same assay, we determined whether Ro 48-8071 would affect normal mammary cells. Our data showed that Ro 48-8071 exhibits significantly less inhibition of normal mammary cells from line AG11132A (Figure 5.4(c)), indicating an effect that is specific to tumor cells. From our Western blot analysis, the expression of OSC was confirmed for both cancer cell lines (BT-474 and T47D), and normal mammary cells showed significantly less expression of OSC.

5.3.3 p53 Activation Assay

Finally, it is well known that PRIMA-1 increases the DNA-binding affinity of mutant p53, which is highly expressed in several different types of cancer cells, including BT-474 and T47D cells [200]. Since Ro 48-8071 and PRIMA-1 exhibited similar inhibition of breast cancer cells (as shown above), we examined the capacity of Ro 47-8071 to restore the DNA-binding of mutant p53 in BT-474 cells, by using a TransAM p53 Transcription Factor Assay kit (Active Motif, Carlsbad, CA). In a time-course study we found that treatment of BT-474 cells with either 25 μM Ro 48-8071 or 50 μM PRIMA-1 for 0.5 to 3.0 hours led to the activation of mutant p53 activity (data not shown). Figure 5.5 compares the extent of mutant p53 activation in BT-474 cells following 1-hour exposure to Ro 48-8071 (25 μM) and PRIMA-1 (50 μM). Treatment with either Ro 48-8071 or PRIMA-1 increased the binding of mutant p53 to DNA. MCF-7 (wild-type p53) nuclear extract treated with H2O2 was provided in the TransAM kit, and used as a positive control.

5.4 Conclusions

In this paper, we presented an application of inverse docking using our software package MDock. Our *in silico* screening identified OSC as one possible target of PRIMA-1. This led us to investigate whether the potent OSC-inhibitor Ro 48-8071 would selectively reduce the viability of human breast cancer cells. It does, and in addition leads to increased binding of mutant p53 to DNA. These effects of Ro 48-8071 are similar to the corresponding characteristic effects of PRIMA-1. In conjunction with our computational mechanistic study, these results lead us to suspect that these two ligands are exerting their anti-cancer effects in part due to inhibition of OSC, but it remains to be shown experimentally whether or not PRIMA-1, like Ro 48-8071, binds directly to OSC. Given the potent inhibition of breast cancer cells induced by Ro 48-8071, we consider it and other OSC inhibitors worth investigating as possible

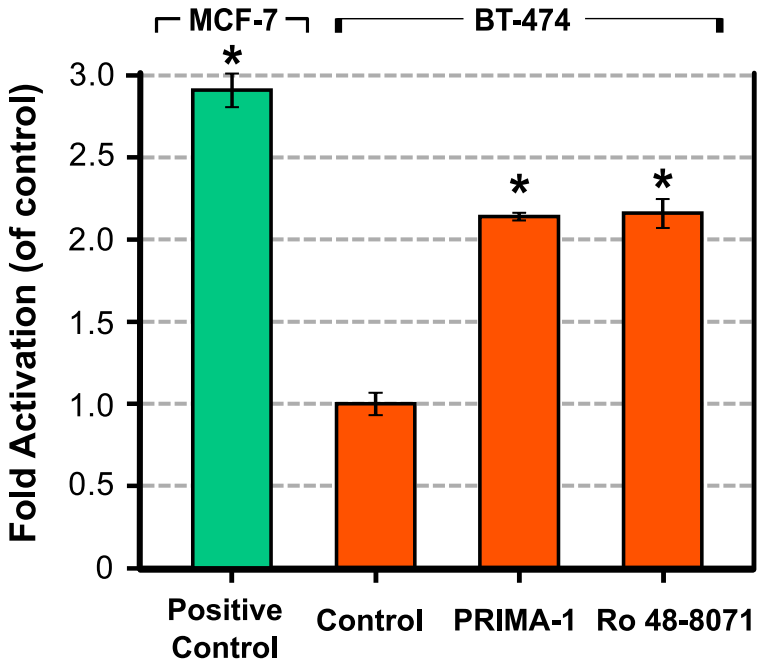

Figure 5.5: Both PRIMA-1 and Ro 48-8071 increase p53-DNA binding in BT-474 breast cancer cells. BT-474 cells were grown in DMEM/F12 medium supplemented with 5% FBS overnight. Cells were then washed with PBS once and treated with 50 µM PRIMA-1 or 25 µM Ro 48-8071 for 1 hour. Cells were harvested by scraping and nuclear extracts were prepared. 2.5 µg of nuclear extract were used for each TransAM assay and each sample was analyzed in triplicate. The fold of activation was compared to the control group (i.e., without PRIMA-1 or Ro 48-8071 treatment). MCF-7 nuclear extract treated with H2O2 provided by the TransAM kit was used as a positive control. Data are shown as the Mean ± SEM from three different determinations. Asterisk represents values differing significantly from the untreated BT-474 control ($P < 0.05$). (Data provided by co-corresponding author Dr. Salman M. Hyder and co-first author Dr. Yayun Liang.)

therapeutic agents against breast cancer. The present study is an onset of a series of future experimental and theoretical studies exploring OSC as a new potential target for developing anticancer therapies. Other future studies include conducting the direct binding assay of OSC for PRIMA-1 and testing the inhibitors of other proteins in the top list of our inverse docking study for their ability to inhibit cancer cell growth.

APPENDIX A

CSV Data File Specification

We have prepared a CSV data file and several scripts which we believe will be of use to future users of the CSAR Dataset. The CSV datafile, which is available in the Supporting Information mentioned at the end of Chapter 4, specifies several labels for each of the 757 compounds in the CSAR Dataset. By providing the information for each compound in one file, it becomes simple to apply commands to relevant subsets of the CSAR Dataset. For copyright reasons, the affinity columns from the CSV file have been excluded. The full datafile, including all affinity data, is available on the CSAR website (http://www.csardock.org). The set of scripts we used to set up the CSAR Dataset for docking (*i.e.* the steps in the Methods before docking) are also available at the same website. One of these scripts can be used to generate a set of ligand conformation files with canonicalized atom orders, and may be adapted for other docking applications. The CSV file is described in detail as follows.

We assigned each of the 757 compounds in the CSAR Dataset a consistent filename of the form *code_id* where *code* specifies the protein and *id* specifies the compound ID number. There are six protein groups: CDK2-Cyclin A, CDK2, CHK1, ERK2, LPXC, and Urokinase. The corresponding three-letter codes are CDC, CDK, CHK, ERK, LPX, and URO, respectively. For each compound, we built a CSV file which combines all the basic information available for each compound (other than its structural coordinates) into one file. The data series provided in this CSV file are summarized in Table A.1. The first column gives the compound ID and the second column gives its corresponding protein group. The compound IDs are unique only within each protein group. The third column, the base filename for each compound, combines the compound ID number and protein group name. These base filenames are

Column	Label	Value
1	Compound ID	integer
2	Protein Code	string (CDC, CDK, CHK, ERK, LPX or URO)
3	Base Filename	string (e.g. CDK_4)
4	Excluded?	boolean (0 or 1)
5	Designated active?	boolean (0 or 1)
6	Designated inactive?	boolean (0 or 1)
7	Structure available?	boolean (0 or 1)
8	Affinity available?	boolean (0 or 1)
9	In March structure subset?	boolean (0 or 1)
10	Affinity Measure	string (K_d, K_i, or IC_{50})
11	Affinity Assay	string (OctetRed, Abbott, Vertex, or Thermofluor)
12	Affinity in μM	float
13	Affinity in M	float
14	-log(Affinity in M)	float
15	Compound SMILES	string

Table A.1: CSV Datafile Specification. This table defines the 14 columns listed in our CSV datafile for the 2012 CSAR Dataset. Each column represents a compound label; they are explained in detail in the Appendix.

unique.

The next six columns in the CSV file provide binary labels, most of which describe what information is available for each compound. Logical combinations of these may be used to apply commands to relevant subsets. The first of the binary labels specifies whether or not a compound is excluded from the final results calculations. For our calculations, we only excluded one compound, CDK_5, because we had a doubt about its affinity data. In the CSAR Dataset, the affinity data spreadsheet for CDK2 includes a table of designated inactive compounds. For all of these inactives, the affinity is given as $> 100\mu M$. The other table of compounds includes those with known activity from $0.023\mu M$ to $58.3\mu M$ and two additional compounds CDK_5 and CDK_12. The affinity of CDK_12 is unknown, but the affinity of CDK_5 is given as $K_d > 100\mu M$, the same value as given for the compounds in the list of designated inactives. Nevertheless, the K_d value given for this compound was not trustable, because the compound was reported to be insoluable. Therefore, we were suspicious of CDK_5 and excluded it from our results calculations. CDC_5 is the same compound, and

Protein Name	Code	Total	Doubted	Active	Inactive	Structure	Affinity	March
CDK2-Cyclin A	CDC	111	1	22	84	1	23	0
CDK2	CDK	111	1	22	84	15	25	0
CHK1	CHK	159	0	106	0	17	107	14
ERK2	ERK	298	0	293	0	12	298	12
LPXC	LPX	32	0	13	12	5	20	4
Urokinase	URO	46	0	35	5	7	35	4
	TOTALS	757	2	491	185	57	508	34

Table A.2: Size of CSAR Subsets. This table gives the number of compounds in each protein group that satisfy each of the binary labels. These labels are listed in the column headers.

was also excluded.

The next two binary labels (Columns 5 and 6) specify whether a compound is designated to be active or designated to be inactive. We considered a compound to be active only if its affinity was known to be less than $10\mu M$ (whether K_d, K_i, or IC_{50}). Due to the differences between these affinity measures and assays, we left a gap from $10\mu M$ to $100\mu M$ between which a compound is considered to be neither nor inactive. Compounds with K_d, K_i, or $IC_{50} > 100\mu M$ were labeled as inactive.

Column 7 specifies whether or not a crystal structure is available in the CSAR Dataset containing the compound bound to its associated protein. This column defines the structure subset referred to in the Methods. Likewise, Column 8 specifies if the precise affinity of the compound for its associated protein is available in the CSAR Dataset. We labeled this field as 0 (false) when the affinity is given as a comparison (e.g. $> 100\mu M$). Finally, the last binary label (Column 9) specifies if the protein-ligand structure for a compound was one of the structures in the results of the 2011-2012 CSAR Benchmark Exercise, provided in March. The number of compounds in the CSAR Database satisfying each condition is given in Table A.2.

Column 10 specifies which affinity measure was used (K_d, K_i, or IC_{50}) and Column 11 states which assay was used to produce the data (OctetRed, Abbott, Vertex, or Thermofluor). The next three columns are designated for the affinity data itself. Column 12 gives the affinity of the compound for its associated protein in units of μM, Column 13 gives the affinity in

units of *M*, and Column 14 gives the negative logarithm of the affinity. For three compounds, the original CSAR Dataset includes duplicate entries with slightly different affinities. These duplicate entries represent different salt forms of the same compound. For these compounds, we provide the mean of the two values. Lastly we included all the SMILES strings in the CSAR Dataset in Column 15. The version of this file provided in the Supporting Information excludes the affinity data and SMILES strings. A version providing all data can be found on CSAR's website (http://www.csardock.org/).

Bibliography

[1] Tanaka, S.; Scheraga, H.A. Model of protein folding: incorporation of a one-dimensional short-range (Ising) model into a three-dimensional model. *Proc. Natl. Acad. Sci. U S A* **1977**, *74*, 1320–1323.

[2] Thomas, P.D.; Dill, K.A. An iterative method for extracting energy-like quantities from protein structures. *Proc. Natl. Acad. Sci. U S A* **1996**, *93*, 11628–11633.

[3] Thomas, P.D.; Dill, K.A. Statistical potentials extracted from protein structures: how accurate are they? *J. Mol. Biol.* **1996**, *257*, 457–469.

[4] Muegge, I.; Martin, Y.C.; Hajduk, P.J.; Fesik, S.W. Evaluation of PMF scoring in docking weak ligands to the FK506 binding protein. *J. Med. Chem.* **1999**, *42*, 2498–2503.

[5] Huang, S.Y.; Zou, X. Chapter 14 - Mean-Force Scoring Functions for Protein–Ligand Binding. In *Annual Reports in Computational Chemistry*; Ralph A. Wheeler., Ed.; Elsevier, 2010; Vol. Volume 6, *Annual Reports in Computational Chemistry*, pp. 280–296.

[6] Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287–2303.

[7] Dunbar, James B, J.; Smith, R.D.; Damm-Ganamet, K.L.; Ahmed, A.; Esposito, E.X.; Delproposto, J.; Chinnaswamy, K.; Kang, Y.N.; Kubish, G.; Gestwicki, J.E.; Stuckey, J.A.; Carlson, H.A. CSAR data set release 2012: ligands, affinities, complexes, and docking decoys. *J. Chem. Inf. Model.* **2013**, *53*, 1842–1852.

[8] Bykov, V.J.N.; Issaeva, N.; Shilov, A.; Hultcrantz, M.; Pugacheva, E.; Chumakov, P.; Bergman, J.; Wiman, K.G.; Selivanova, G. Restoration of the tumor suppres-

sor function to mutant p53 by a low-molecular-weight compound. *Nat. Med.* **2002**, *8*, 282–288.

[9] Korb, O.; Olsson, T.S.G.; Bowden, S.J.; Hall, R.J.; Verdonk, M.L.; Liebeschuetz, J.W.; Cole, J.C. Potential and limitations of ensemble docking. *J. Chem. Inf. Model.* **2012**, *52*, 1262–1274.

[10] Liljefors, T.; Krogsgaard-Larsen, P.; Madsen, U. *Textbook of Drug Design and Discovery, Third Edition*; CRC Press, 2003.

[11] Khanna, I. Drug discovery in pharmaceutical industry: productivity challenges and trends. *Drug Discov. Today* **2012**, *17*, 1088–1102.

[12] Overington, J.P.; Al-Lazikani, B.; Hopkins, A.L. How many drug targets are there? *Nat. Rev. Drug Discov.* **2006**, *5*, 993–996.

[13] Feher, M.; Schmidt, J.M. Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 218–227.

[14] Landon, M.R.; Lancia, David R, J.; Yu, J.; Thiel, S.C.; Vajda, S. Identification of hot spots within druggable binding regions by computational solvent mapping of proteins. *J. Med. Chem.* **2007**, *50*, 1231–1240.

[15] Brenke, R.; Kozakov, D.; Chuang, G.Y.; Beglov, D.; Hall, D.; Landon, M.R.; Mattos, C.; Vajda, S. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **2009**, *25*, 621–627.

[16] Schneider, G.; Böhm, H.J. Virtual screening and fast automated docking methods. *Drug Discov. Today* **2002**, *7*, 64–70.

[17] Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J.L.; Martínez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D.K. Recognizing pitfalls in virtual screening: a critical review. *J. Chem. Inf. Model.* **2012**, *52*, 867–881.

[18] Irwin, J.J.; Shoichet, B.K. ZINC–a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.

[19] Brown, R.D.; Martin, Y.C. The Information Content of 2D and 3D Structural Descriptors Relevant to Ligand-Receptor Binding. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1–9.

[20] Patani, G.A.; LaVoie, E.J. Bioisosterism: A Rational Approach in Drug Design. *Chem. Rev.* **1996**, *96*, 3147–3176.

[21] Hansch, C. The physicochemical approach to drug design and discovery (QSAR). *Drug Dev. Res.* **1981**, *1*, 267–309.

[22] Horvath, D. Pharmacophore-based virtual screening. *Methods Mol. Biol.* **2011**, *672*, 261–298.

[23] Willett, P. Similarity searching using 2D structural fingerprints. *Methods Mol. Biol.* **2011**, *672*, 133–158.

[24] Durant, J.L.; Leland, B.A.; Henry, D.R.; Nourse, J.G. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273–1280.

[25] Radifar, M.; Yuniarti, N.; Istyastono, E.P. PyPLIF: Python-based Protein-Ligand Interaction Fingerprinting. *Bioinformation* **2013**, *9*, 325–328.

[26] Lyne, P.D. Structure-based virtual screening: an overview. *Drug Discov. Today* **2002**, *7*, 1047–1055.

[27] Brooijmans, N.; Kuntz, I.D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373.

[28] Leach, A.R.; Shoichet, B.K.; Peishoff, C.E. Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. *J. Med. Chem.* **2006**, *49*, 5851–5855.

[29] Huang, S.Y.; Zou, X. Advances and Challenges in Protein-Ligand Docking. *Int. J. Mol. Sci.* **2010**, *11*, 3016–3034.

[30] Cereto-Massagué, A.; Ojeda, M.J.; Joosten, R.P.; Valls, C.; Mulero, M.; Salvado, M.J.; Arola-Arnal, A.; Arola, L.; Garcia-Vallvé, S.; Pujadas, G. The good, the bad and the dubious: VHELIBS, a validation helper for ligands and binding sites. *J. Cheminform.* **2013**, *5*, 36.

[31] DesJarlais, R.L.; Sheridan, R.P.; Seibel, G.L.; Dixon, J.S.; Kuntz, I.D.; Venkataraghavan, R. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J. Med. Chem.* **1988**, *31*, 722–729.

[32] Meng, E.C.; Shoichet, B.K.; Kuntz, I.D. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* **1992**, *13*, 505–524.

[33] Kuntz, I.D.; Meng, E.C.; Shoichet, B.K. Structure-Based Molecular Design. *Acc. Chem. Res.* **1994**, *27*, 117–123.

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

[35] Moustakas, D.T.; Lang, P.T.; Pegg, S.; Pettersen, E.; Kuntz, I.D.; Brooijmans, N.; Rizzo, R.C. Development and validation of a modular, extensible docking program: DOCK 5. *J. Comput. Aided Mol. Des.* **2006**, *20*, 601–619.

[36] Lang, P.T.; Brozell, S.R.; Mukherjee, S.; Pettersen, E.F.; Meng, E.C.; Thomas, V.; Rizzo,

R.C.; Case, D.A.; James, T.L.; Kuntz, I.D. DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA* **2009**, *15*, 1219–1230.

[37] Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.

[38] Böhm, H.J. The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J. Comput. Aided Mol. Des.* **1992**, *6*, 61–78.

[39] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.

[40] Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* **1999**, *37*, 228–241.

[41] Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.

[42] Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; Shaw, D.E.; Francis, P.; Shenkin, P.S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.

[43] Halgren, T.A.; Murphy, R.B.; Friesner, R.A.; Beard, H.S.; Frye, L.L.; Pollard, W.T.; Banks, J.L. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.

[44] Trott, O.; Olson, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461.

[45] Huang, S.Y.; Zou, X. An iterative knowledge-based scoring function to predict protein-ligand interactions: I. Derivation of interaction potentials. *J. Comput. Chem.* **2006**, *27*, 1866–1875.

[46] Huang, S.Y.; Zou, X. An iterative knowledge-based scoring function to predict protein-ligand interactions: II. Validation of the scoring function. *J. Comput. Chem.* **2006**, *27*, 1876–1882.

[47] Huang, S.Y.; Zou, X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. *Proteins* **2007**, *66*, 399–421.

[48] Huang, S.Y.; Zou, X. Efficient molecular docking of NMR structures: application to HIV-1 protease. *Protein Sci.* **2007**, *16*, 43–51.

[49] Huang, S.Y.; Zou, X. Inclusion of solvation and entropy in the knowledge-based scoring function for protein-ligand interactions. *J. Chem. Inf. Model.* **2010**, *50*, 262–273.

[50] Sousa, S.F.; Ribeiro, A.J.M.; Coimbra, J.T.S.; Neves, R.P.P.; Martins, S.A.; Moorthy, N.S.H.N.; Fernandes, P.A.; Ramos, M.J. Protein-ligand docking in the new millennium–a retrospective of 10 years in the field. *Curr. Med. Chem.* **2013**, *20*, 2296–2314.

[51] Zhou, H.; Skolnick, J. FINDSITE(comb): a threading/structure-based, proteomic-scale virtual ligand screening approach. *J. Chem. Inf. Model.* **2013**, *53*, 230–240.

[52] Anderson, A.C. The process of structure-based drug design. *Chem. Biol.* **2003**, *10*, 787–797.

[53] Gilson, M.K.; Zhou, H.X. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.

[54] Huang, S.Y.; Grinter, S.Z.; Zou, X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12899–12908.

[55] Kitchen, D.B.; Decornez, H.; Furr, J.R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **2004**, *3*, 935–949.

[56] Rahaman, O.; Estrada, T.P.; Doren, D.J.; Taufer, M.; Brooks, Charles L, r.; Armen, R.S. Evaluation of several two-step scoring functions based on linear interaction energy, effective ligand size, and empirical pair potentials for prediction of protein-ligand binding geometry and free energy. *J. Chem. Inf. Model.* **2011**, *51*, 2047–2065.

[57] Nicolini, P.; Frezzato, D.; Gellini, C.; Bizzarri, M.; Chelli, R. Toward quantitative estimates of binding affinities for protein-ligand systems involving large inhibitor compounds: a steered molecular dynamics simulation route. *J. Comput. Chem.* **2013**, *34*, 1561–1576.

[58] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712–725.

[59] Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

[60] Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

[61] Jones, J.E. On the Determination of Molecular Fields. II. From the Equation of State of a Gas. *Proc. R. Soc. Lond. A* **1924**, *106*, 463–477.

[62] MacKerell, A.D.; Bashford, D.; Bellott.; Dunbrack, R.L.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.T.K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.T.; Prodhom, B.; Reiher, W.E.; Roux, B.; Schlenkrich, M.; Smith, J.C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

[63] Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A D, J. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671–690.

[64] Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

[65] Gilson, M.K.; Rashin, A.; Fine, R.; Honig, B. On the calculation of electrostatic interactions in proteins. *J. Mol. Biol.* **1985**, *184*, 503–516.

[66] Grant, J.A.; Pickup, B.T.; Nicholls, A. A smooth permittivity function for Poisson–Boltzmann solvation methods. *J. Comput. Chem.* **2001**, *22*, 608–640.

[67] Baker, N.A.; Sept, D.; Joseph, S.; Holst, M.J.; McCammon, J.A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.

[68] Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. Rapid

grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J. Comput. Chem.* **2002**, *23*, 128–137.

[69] Still, W.C.; Tempczyk, A.; Hawley, R.C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

[70] Bashford, D.; Case, D.A. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–152.

[71] Hawkins, G.D.; Cramer, C.J.; Truhlar, D.G. Pairwise solute descreening of solute charges from a dielectric medium. *Chemical Physics Letters* **1995**, *246*, 122–129.

[72] Grycuk, T. Deficiency of the Coulomb-field approximation in the generalized Born model: An improved formula for Born radii evaluation. *The Journal of Chemical Physics* **2003**, *119*, 4817–4826.

[73] Feig, M.; Onufriev, A.; Lee, M.S.; Im, W.; Case, D.A.; Brooks, Charles L, r. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* **2004**, *25*, 265–284.

[74] Liu, H.Y.; Zou, X. Electrostatics of ligand binding: parametrization of the generalized Born model and comparison with the Poisson-Boltzmann approach. *J. Phys. Chem. B* **2006**, *110*, 9304–9313.

[75] Tjong, H.; Zhou, H.X. GBr(6): a parameterization-free, accurate, analytical generalized born method. *J. Phys. Chem. B* **2007**, *111*, 3055–3061.

[76] Srinivasan, J.; Miller, J.; Kollman, P.A.; Case, D.A. Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dyn.* **1998**, *16*, 671–682.

[77] Zou, X.; Yaxiong.; Kuntz, I.D. Inclusion of Solvation in Ligand Binding Free Energy Calculations Using the Generalized-Born Model. *J. Am. Chem. Soc.* **1999**, *121*, 8033–8043.

[78] Wang, J.; Morin, P.; Wang, W.; Kollman, P.A. Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA. *J. Am. Chem. Soc.* **2001**, *123*, 5221–5230.

[79] Zhou, R. Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins* **2003**, *53*, 148–161.

[80] Liu, H.Y.; Kuntz, I.D.; Zou, X. Pairwise GB/SA Scoring Function for Structure-based Drug Design. *J. Phys. Chem. B* **2004**, *108*, 5453–5462.

[81] Liu, H.Y.; Grinter, S.Z.; Zou, X. Multiscale generalized Born modeling of ligand binding energies for virtual database screening. *J. Phys. Chem. B* **2009**, *113*, 11793–11799.

[82] Purisima, E.O.; Hogues, H. Protein-ligand binding free energies from exhaustive docking. *J. Phys. Chem. B* **2012**, *116*, 6872–6879.

[83] Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.

[84] Eldridge, M.D.; Murray, C.W.; Auton, T.R.; Paolini, G.V.; Mee, R.P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the

binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.* **1997**, *11*, 425–445.

[85] Böhm, H.J. Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput. Aided Mol. Des.* **1998**, *12*, 309–323.

[86] Temiz, N.A.; Trapp, A.; Prokopyev, O.A.; Camacho, C.J. Optimization of minimum set of protein-DNA interactions: a quasi exact solution with minimum over-fitting. *Bioinformatics* **2010**, *26*, 319–325.

[87] Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.* **2002**, *16*, 11–26.

[88] Miyazawa, S.; Jernigan, R.L. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* **1985**, *18*, 534–552.

[89] Sippl, M.J.; Ortner, M.; Jaritz, M.; Lackner, P.; Flöckner, H. Helmholtz free energies of atom pair interactions in proteins. *Fold Des.* **1996**, *1*, 289–298.

[90] Li, X.; Liang, J. Knowledge-based energy functions for computational studies of proteins. In *Computational Methods for Protein Structure Prediction and Modeling*; Springer New York, 2007; pp. 71–123.

[91] Munson, P.J.; Singh, R.K. Statistical significance of hierarchical multi-body potentials based on Delaunay tessellation and their application in sequence-structure alignment. *Protein Sci.* **1997**, *6*, 1467–1481.

[92] Zimmermann, M.T.; Leelananda, S.P.; Kloczkowski, A.; Jernigan, R.L. Combining statistical potentials with dynamics-based entropies improves selection from protein decoys and docking poses. *J. Phys. Chem. B* **2012**, *116*, 6725–6731.

[93] Jernigan, R.L.; Bahar, I. Structure-derived potentials and protein simulations. *Curr. Opin. Struct. Biol.* **1996**, *6*, 195–209.

[94] Zhang, L.; Skolnick, J. How do potentials derived from structural databases relate to "true" potentials? *Protein Sci.* **1998**, *7*, 112–122.

[95] Muegge, I.; Martin, Y.C. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791–804.

[96] Zhou, H.; Zhou, Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci.* **2002**, *11*, 2714–2726.

[97] Kozakov, D.; Brenke, R.; Comeau, S.R.; Vajda, S. PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins* **2006**, *65*, 392–406.

[98] Huang, S.Y.; Zou, X. An iterative knowledge-based scoring function for protein-protein recognition. *Proteins* **2008**, *72*, 557–579.

[99] Ravikant, D.V.S.; Elber, R. Energy design for protein-protein interactions. *J. Chem. Phys.* **2011**, *135*, 065102.

[100] Huang, S.Y.; Zou, X. A knowledge-based scoring function for protein-RNA interactions derived from a statistical mechanics-based iterative method. *Nucleic Acids Res.* **2014**.

[101] Sippl, M.J. Calculation of conformational ensembles from potentials of mean force. An

approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.* **1990**, *213*, 859–883.

[102] Grinter, S.Z.; Zou, X. A Bayesian statistical approach of improving knowledge-based scoring functions for protein-ligand interactions. *J. Comput. Chem.* **2014**, *35*, 932–943.

[103] Zhang, C.; Liu, S.; Zhu, Q.; Zhou, Y. A knowledge-based energy function for protein-ligand, protein-protein, and protein-DNA complexes. *J. Med. Chem.* **2005**, *48*, 2325–2335.

[104] Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.

[105] Velec, H.F.G.; Gohlke, H.; Klebe, G. DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.* **2005**, *48*, 6296–6303.

[106] DeWitte, R.; Shakhnovich, E. SMoG: de Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence. *J. Am. Chem. Soc.* **1996**, *118*, 11733–11744.

[107] Charifson, P.S.; Corkery, J.J.; Murcko, M.A.; Walters, W.P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100–5109.

[108] Terp, G.E.; Johansen, B.N.; Christensen, I.T.; Jørgensen, F.S. A new concept for multidimensional selection of ligand conformations (MultiSelect) and multidimensional scoring (MultiScore) of protein-ligand binding affinities. *J. Med. Chem.* **2001**, *44*, 2333–2343.

[109] Plewczynski, D.; Łazniewski, M.; von Grotthuss, M.; Rychlewski, L.; Ginalski, K. VoteDock: consensus docking method for prediction of protein-ligand interactions. *J. Comput. Chem.* **2011**, *32*, 568–581.

[110] Erickson, J.A.; Jalaie, M.; Robertson, D.H.; Lewis, R.A.; Vieth, M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* **2004**, *47*, 45–55.

[111] Shoichet, B.K.; Kuntz, I.D.; Bodian, D.L. Molecular docking using shape descriptors. *J. Comput. Chem.* **1992**, *13*, 380–397.

[112] Hawkins, P.C.D.; Skillman, A.G.; Warren, G.L.; Ellingson, B.A.; Stahl, M.T. Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572–584.

[113] Hawkins, P.C.D.; Nicholls, A. Conformer generation with OMEGA: learning from the data set and the analysis of failures. *J. Chem. Inf. Model.* **2012**, *52*, 2919–2936.

[114] Leach, A.R.; Kuntz, I.D. Conformational analysis of flexible ligands in macromolecular receptor sites. *Journal of Computational Chemistry* **1992**, *13*, 730–748.

[115] Lorber, D.M.; Shoichet, B.K. Hierarchical docking of databases of multiple ligand conformations. *Curr. Top. Med. Chem.* **2005**, *5*, 739–749.

[116] Damm, K.L.; Carlson, H.A. Exploring experimental sources of multiple protein conformations in structure-based drug design. *J. Am. Chem. Soc.* **2007**, *129*, 8225–8235.

[117] Bottegoni, G.; Kufareva, I.; Totrov, M.; Abagyan, R. Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking. *J. Med. Chem.* **2009**, *52*, 397–406.

[118] Apostolakis, J.; Plückthun, A.; Caflisch, A. Docking small ligands in flexible binding sites. *J. Comput. Chem.* **1998**, *19*, 21–37.

[119] Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791.

[120] Limongelli, V.; Marinelli, L.; Cosconati, S.; La Motta, C.; Sartini, S.; Mugnaini, L.; Da Settimo, F.; Novellino, E.; Parrinello, M. Sampling protein motion and solvent effect during ligand binding. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 1467–1472.

[121] Rarey, M.; Kramer, B.; Lengauer, T. The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins* **1999**, *34*, 17–28.

[122] Sahai, M.A.; Biggin, P.C. Quantifying water-mediated protein-ligand interactions in a glutamate receptor: a DFT study. *J. Phys. Chem. B* **2011**, *115*, 7085–7096.

[123] Lie, M.A.; Thomsen, R.; Pedersen, C.N.S.; Schiøtt, B.; Christensen, M.H. Molecular docking with ligand attached water molecules. *J. Chem. Inf. Model.* **2011**, *51*, 909–917.

[124] Liu, J.; He, X.; Zhang, J.Z.H. Improving the scoring of protein-ligand binding affinity by including the effects of structural water and electronic polarization. *J. Chem. Inf. Model.* **2013**, *53*, 1306–1314.

[125] Wang, C.; Bradley, P.; Baker, D. Protein-protein docking with backbone flexibility. *J. Mol. Biol.* **2007**, *373*, 503–519.

[126] Lemmon, G.; Meiler, J. Rosetta Ligand docking with flexible XML protocols. *Methods Mol. Biol.* **2012**, *819*, 143–155.

[127] Huggins, D.J.; Tidor, B. Systematic placement of structural water molecules for improved scoring of protein-ligand interactions. *Protein Eng. Des. Sel.* **2011**, *24*, 777–789.

[128] Reddy, A.S.; Zhang, S. Polypharmacology: drug discovery for the future. *Expert Rev. Clin. Pharmacol.* **2013**, *6*, 41–47.

[129] Taboureau, O.; Jørgensen, F.S. In silico predictions of hERG channel blockers in drug discovery: from ligand-based and target-based approaches to systems chemical biology. *Comb. Chem. High Throughput Screen.* **2011**, *14*, 375–387.

[130] Gowthaman, R.; Deeds, E.J.; Karanicolas, J. Structural properties of non-traditional drug targets present new challenges for virtual screening. *J. Chem. Inf. Model.* **2013**, *53*, 2073–2081.

[131] Pérez-Nueno, V.I.; Ritchie, D.W. Using consensus-shape clustering to identify promiscuous ligands and protein targets and to choose the right query for shape-based virtual screening. *J. Chem. Inf. Model.* **2011**, *51*, 1233–1248.

[132] Peng, S.; Lin, X.; Guo, Z.; Huang, N. Identifying multiple-target ligands via computational chemogenomics approaches. *Curr. Top. Med. Chem.* **2012**, *12*, 1363–1375.

[133] Shrinivasan, M.; Skariyachan, S.; Aparna, V.; Kolte, V.R. Homology modelling of CB1 receptor and selection of potential inhibitor against Obesity. *Bioinformation* **2012**, *8*, 523–528.

[134] Skariyachan, S.; Mahajanakatti, A.B.; Sharma, N.; Karanth, S.; Rao, S.; Rajeswari, N. Structure based virtual screening of novel inhibitors against multidrug resistant superbugs. *Bioinformation* **2012**, *8*, 420–425.

[135] Skariyachan, S.; Prakash, N.; Bharadwaj, N. In silico exploration of novel phytoligands against probable drug target of Clostridium tetani. *Interdiscip. Sci.* **2012**, *4*, 273–281.

[136] Kar, R.K.; Ansari, M.Y.; Suryadevara, P.; Sahoo, B.R.; Sahoo, G.C.; Dikhit, M.R.; Das, P. Computational elucidation of structural basis for ligand binding with Leishmania donovani adenosine kinase. *Biomed. Res. Int.* **2013**, *2013*, 609289.

[137] Tahir, R.A.; Sehgal, S.A.; Khattak, N.A.; Khan Khattak, J.Z.; Mir, A. Tumor necrosis factor receptor superfamily 10B (TNFRSF10B): an insight from structure modeling to virtual screening for designing drug against head and neck cancer. *Theor. Biol. Med. Model.* **2013**, *10*, 38.

[138] Skariyachan, S.; Jayaprakash, N.; Bharadwaj, N.; Narayanappa, R. Exploring insights for virulent gene inhibition of multidrug resistant Salmonella typhi, Vibrio cholerae, and Staphylococcus areus by potential phytoligands via in silico screeningd. *J. Biomol. Struct. Dyn.* **2013**.

[139] Merlino, A.; Vieites, M.; Gambino, D.; Laura Coitiño, E. Homology modeling of T. cruzi and L. major NADH-dependent fumarate reductases: Ligand docking, molecular dynamics validation, and insights on their binding modes. *J. Mol. Graph. Model.* **2014**, *48*, 47–59.

[140] Orry, A.J.W.; Abagyan, R. Preparation and refinement of model protein-ligand complexes. *Methods Mol. Biol.* **2012**, *857*, 351–373.

[141] Combs, S.A.; Deluca, S.L.; Deluca, S.H.; Lemmon, G.H.; Nannemann, D.P.; Nguyen, E.D.; Willis, J.R.; Sheehan, J.H.; Meiler, J. Small-molecule ligand docking into comparative models with Rosetta. *Nat. Protoc.* **2013**, *8*, 1277–1298.

[142] Kaufmann, K.W.; Meiler, J. Using RosettaLigand for small molecule docking into comparative models. *PLoS ONE* **2012**, *7*, e50769.

[143] Mahasenan, K.V.; Li, C. Novel inhibitor discovery through virtual screening against multiple protein conformations generated via ligand-directed modeling: a maternal embryonic leucine zipper kinase example. *J. Chem. Inf. Model.* **2012**, *52*, 1345–1355.

[144] Sali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.

[145] Kiss, R.; Sandor, M.; Szalai, F.A. http://Mcule.com: a public web service for drug discovery. *J. Cheminform.* **2012**, *4*, P17.

[146] Zhang, X.; Wang, H.; Li, Y.; Cao, R.; Zhong, W.; Zheng, Z.; Wang, G.; Xiao, J.; Li, S. Novel substituted heteroaromatic piperazine and piperidine derivatives as inhibitors of human enterovirus 71 and coxsackievirus a16. *Molecules* **2013**, *18*, 5059–5071.

[147] Wilson, G.L.; Lill, M.A. Integrating structure-based and ligand-based approaches for computational drug design. *Future Med. Chem.* **2011**, *3*, 735–750.

[148] Ahmed, L.; Rasulev, B.; Turabekova, M.; Leszczynska, D.; Leszczynski, J. Receptor- and ligand-based study of fullerene analogues: comprehensive computational approach including quantum-chemical, QSAR and molecular docking simulations. *Org. Biomol. Chem.* **2013**, *11*, 5798–5808.

[149] Ballante, F.; Caroli, A.; Wickersham, Richard B, r.; Ragno, R. Hsp90 Inhibitors, Part 1: Definition of 3-D QSAutogrid/R Models as a Tool for Virtual Screening. *J. Chem. Inf. Model.* **2014**, *54*, 956–969.

[150] Caroli, A.; Ballante, F.; Wickersham, Richard B, r.; Corelli, F.; Ragno, R. Hsp90 In-

hibitors, Part 2: Combining Ligand-Based and Structure-Based Approaches for Virtual Screening Application. *J. Chem. Inf. Model.* **2014**, *54*, 970–977.

[151] Alcaro, S.; Musetti, C.; Distinto, S.; Casatti, M.; Zagotto, G.; Artese, A.; Parrotta, L.; Moraca, F.; Costa, G.; Ortuso, F.; Maccioni, E.; Sissi, C. Identification and characterization of new DNA G-quadruplex binders selected by a combination of ligand and structure-based virtual screening approaches. *J. Med. Chem.* **2013**, *56*, 843–855.

[152] Grinter, S.Z.; Liang, Y.; Huang, S.Y.; Hyder, S.M.; Zou, X. An inverse docking approach for identifying new potential anti-cancer targets. *J. Mol. Graph. Model.* **2011**, *29*, 795–799.

[153] Chen, Y.Z.; Zhi, D.G. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins* **2001**, *43*, 217–226.

[154] Paul, N.; Kellenberger, E.; Bret, G.; Müller, P.; Rognan, D. Recovering the true targets of specific ligands by virtual screening of the protein data bank. *Proteins* **2004**, *54*, 671–680.

[155] Gao, Z.; Li, H.; Zhang, H.; Liu, X.; Kang, L.; Luo, X.; Zhu, W.; Chen, K.; Wang, X.; Jiang, H. PDTD: a web-accessible protein database for drug target identification. *BMC Bioinformatics* **2008**, *9*, 104.

[156] Kumar, S.P.; Pandya, H.A.; Desai, V.H.; Jasrai, Y.T. Compound prioritization from inverse docking experiment using receptor-centric and ligand-centric methods: a case study on Plasmodium falciparum Fab enzymes. *J. Mol. Recognit.* **2014**, *27*, 215–229.

[157] Ogungbe, I.V.; Setzer, W.N. In-silico Leishmania target selectivity of antiparasitic terpenoids. *Molecules* **2013**, *18*, 7761–7847.

[158] Sakkiah, S.; Arullaperumal, V.; Hwang, S.; Lee, K.W. Ligand-based pharmacophore modeling and Bayesian approaches to identify c-Src inhibitors. *J. Enzyme Inhib. Med. Chem.* **2014**, *29*, 69–80.

[159] Yang, Z.; Yang, G.; Zhou, L. Mutation effects of neuraminidases and their docking with ligands: a molecular dynamics and free energy calculation study. *J. Comput. Aided Mol. Des.* **2013**, *27*, 935–950.

[160] Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J. Med. Chem.* **2004**, *47*, 2977–2980.

[161] Huang, S.Y.; Zou, X. Construction and test of ligand decoy sets using MDock: community structure-activity resource benchmarks for binding mode prediction. *J. Chem. Inf. Model.* **2011**, *51*, 2107–2114.

[162] Dunbar, J.B.; Smith, R.D.; Yang, C.Y.; Ung, P.M.U.; Lexa, K.W.; Khazanov, N.A.; Stuckey, J.A.; Wang, S.; Carlson, H.A. CSAR Benchmark Exercise of 2010: Selection of the Protein-Ligand Complexes. *J. Chem. Inf. Model.* **2011**, *51*, 2036–2046.

[163] Kumar, A.; Zhang, K.Y.J. Computational fragment-based screening using RosettaLigand: the SAMPL3 challenge. *J. Comput. Aided Mol. Des.* **2012**, *26*, 603–616.

[164] Skillman, A.G.; Geballe, M.T.; Nicholls, A. SAMPL2 challenge: prediction of solvation energies and tautomer ratios. *J. Comput. Aided Mol. Des.* **2010**, *24*, 257–258.

[165] Grinter, S.Z.; Yan, C.; Huang, S.Y.; Jiang, L.; Zou, X. Automated large-scale file preparation, docking, and scoring: evaluation of ITScore and STScore using the 2012 Community Structure-Activity Resource benchmark. *J. Chem. Inf. Model.* **2013**, *53*, 1905–1914.

[166] Bolia, A.; Gerek, Z.N.; Ozkan, S.B. BP-Dock: A Flexible Docking Scheme for Exploring Protein-Ligand Interactions Based on Unbound Structures. *J. Chem. Inf. Model.* **2014**, *54*, 913–925.

[167] Korb, O.; Ten Brink, T.; Victor Paul Raj, F.R.D.; Keil, M.; Exner, T.E. Are predefined decoy sets of ligand poses able to quantify scoring function accuracy? *J. Comput. Aided Mol. Des.* **2012**, *26*, 185–197.

[168] Vajda, S.; Hall, D.R.; Kozakov, D. Sampling and scoring: a marriage made in heaven. *Proteins* **2013**, *81*, 1874–1884.

[169] Allen, W.J.; Rizzo, R.C. Implementation of the hungarian algorithm to account for ligand symmetry and similarity in structure-based design. *J. Chem. Inf. Model.* **2014**, *54*, 518–529.

[170] Head, M.S.; Given, J.A.; Gilson, M.K. "Mining Minima": Direct Computation of Conformational Free Energy. *J. Phys. Chem. A* **1997**, *101*, 1609–1618.

[171] Ruvinsky, A.M. Role of binding entropy in the refinement of protein-ligand docking predictions: analysis based on the use of 11 scoring functions. *J. Comput. Chem.* **2007**, *28*, 1364–1372.

[172] Tanaka, S.; Scheraga, H.A. Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules* **1976**, *9*, 945–950.

[173] Wang, W.; Donini, O.; Reyes, C.M.; Kollman, P.A. Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211–243.

[174] Mitchell, J.B.O.; Laskowski, R.A.; Alex, A.; Thornton, J.M. BLEEP—potential of mean force describing protein–ligand interactions: I. Generating potential. *J. Comput. Chem.* **1999**, *20*, 1165–1176.

[175] Mitchell, J.B.O.; Laskowski, R.A.; Alex, A.; Forster, M.J.; Thornton, J.M. BLEEP—potential of mean force describing protein–ligand interactions: II. Calculation of binding energies and comparison with experimental data. *J. Comput. Chem.* **1999**, *20*, 1177–1185.

[176] Yang, C.Y.; Wang, R.; Wang, S. M-score: a knowledge-based potential scoring function accounting for protein atom mobility. *J. Med. Chem.* **2006**, *49*, 5903–5911.

[177] Fan, H.; Schneidman-Duhovny, D.; Irwin, J.J.; Dong, G.; Shoichet, B.K.; Sali, A. Statistical potential for modeling and ranking of protein-ligand interactions. *J. Chem. Inf. Model.* **2011**, *51*, 3078–3092.

[178] Muegge, I. A knowledge-based scoring function for protein-ligand interactions: Probing the reference state. *perspect drug discov* **2000**, *20*, 99–114.

[179] Durbin, R. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*; Cambridge University Press, 1998.

[180] Tatusov, R.L.; Altschul, S.F.; Koonin, E.V. Detection of conserved segments in proteins: iterative scanning of sequence databases with alignment blocks. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12091–12095.

[181] Xu, B.; Yang, Y.; Liang, H.; Zhou, Y. An all-atom knowledge-based energy function for protein-DNA threading, docking decoy discrimination, and prediction of transcription-factor binding profiles. *Proteins* **2009**, *76*, 718–730.

[182] Bauer, A.; Beyer, A. An improved pair potential to recognize native protein folds. *Proteins* **1994**, *18*, 254–261.

[183] Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.

[184] Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **1986**, *7*, 230–252.

[185] Wang, R.; Fang, X.; Lu, Y.; Yang, C.Y.; Wang, S. The PDBbind database: methodologies and updates. *J. Med. Chem.* **2005**, *48*, 4111–4119.

[186] Guddat, L.W.; Shan, L.; Anchin, J.M.; Linthicum, D.S.; Edmundson, A.B. Local and transmitted conformational changes on complexation of an anti-sweetener Fab. *J. Mol. Biol.* **1994**, *236*, 247–274.

[187] Huang, N.; Jacobson, M.P. Physics-based methods for studying protein-ligand interactions. *Curr. Opin. Drug Discov. Devel.* **2007**, *10*, 325–331.

[188] Smith, R.D.; Dunbar, James B, J.; Ung, P.M.U.; Esposito, E.X.; Yang, C.Y.; Wang, S.; Carlson, H.A. CSAR benchmark exercise of 2010: combined evaluation across all submitted scoring functions. *J. Chem. Inf. Model.* **2011**, *51*, 2115–2131.

[189] Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

[190] Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.

[191] Sing, T.; Sander, O.; Beerenwinkel, N.; Lengauer, T. ROCR: visualizing classifier performance in R. *Bioinformatics* **2005**, *21*, 3940–3941.

[192] Kyte, J.; Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132.

[193] Huang, S.Y.; Zou, X. Scoring and lessons learned with the CSAR benchmark using an improved iterative knowledge-based scoring function. *J. Chem. Inf. Model.* **2011**, *51*, 2097–2106.

[194] Do, Q.T.; Renimel, I.; Andre, P.; Lugnier, C.; Muller, C.D.; Bernard, P. Reverse pharmacognosy: application of selnergy, a new tool for lead discovery. The example of epsilon-viniferin. *Curr. Drug Discov. Technol.* **2005**, *2*, 161–167.

[195] Muller, P.; Lena, G.; Boilard, E.; Bezzine, S.; Lambeau, G.; Guichard, G.; Rognan, D. In silico-guided target identification of a scaffold-focused library: 1,3,5-triazepan-2,6-diones as novel phospholipase A2 inhibitors. *J. Med. Chem.* **2006**, *49*, 6768–6778.

[196] Zahler, S.; Tietze, S.; Totzke, F.; Kubbutat, M.; Meijer, L.; Vollmar, A.M.; Apostolakis, J. Inverse in silico screening for identification of kinase inhibitor targets. *Chem. Biol.* **2007**, *14*, 1207–1214.

[197] Schapira, M.; Abagyan, R.; Totrov, M. Nuclear hormone receptor targeted virtual screening. *J. Med. Chem.* **2003**, *46*, 3045–3059.

[198] Rollinger, J.M. Accessing target information by virtual parallel screening—The impact on natural product research. *Phytochemistry Letters* **2009**, *2*, 53–58.

[199] Bissantz, C.; Logean, A.; Rognan, D. High-throughput modeling of human G-protein coupled receptors: amino acid sequence alignment, three-dimensional model building, and receptor library screening. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1162–1176.

[200] Liang, Y.; Wu, J.; Stancel, G.M.; Hyder, S.M. p53-dependent inhibition of progestin-induced VEGF expression in human breast cancer cells. *J. Steroid Biochem. Mol. Biol.* **2005**, *93*, 173–182.

[201] Liang, Y.; Besch-Williford, C.; Benakanakere, I.; Hyder, S.M. Re-activation of the p53 pathway inhibits in vivo and in vitro growth of hormone-dependent human breast cancer cells. *Int. J. Oncol.* **2007**, *31*, 777–784.

[202] Liang, Y.; Besch-Williford, C.; Brekken, R.A.; Hyder, S.M. Progestin-dependent progression of human breast tumor xenografts: a novel model for evaluating antitumor therapeutics. *Cancer Res.* **2007**, *67*, 9929–9936.

[203] Wang, T.; Lee, K.; Rehman, A.; Daoud, S.S. PRIMA-1 induces apoptosis by inhibiting JNK signaling but promoting the activation of Bax. *Biochem. Biophys. Res. Commun.* **2007**, *352*, 203–212.

[204] Lambert, J.M.R.; Gorzov, P.; Veprintsev, D.B.; Söderqvist, M.; Segerbäck, D.; Bergman, J.; Fersht, A.R.; Hainaut, P.; Wiman, K.G.; Bykov, V.J.N. PRIMA-1 reactivates mutant p53 by covalent binding to the core domain. *Cancer Cell.* **2009**, *15*, 376–388.

[205] Morand, O.H.; Aebi, J.; Guerry, P.; Hartman, P.G.; Hennes, U.; Himber, J.; Ji, Y.H.; Jolidon, S.; Lengsfeld, H. Potent inhibitors of mammalian 2,3-oxidosqualene lanosterol cyclase are orally active cholesterol lowering agents. *Atherosclerosis* **1994**, *109*, 321.

[206] Lenhart, A.; Reinert, D.J.; Aebi, J.D.; Dehmlow, H.; Morand, O.H.; Schulz, G.E. Binding structures and potencies of oxidosqualene cyclase inhibitors with the homologous squalene-hopene cyclase. *J. Med. Chem.* **2003**, *46*, 2083–2092.

[207] Davidoff, A.M.; Kerns, B.J.; Pence, J.C.; Marks, J.R.; Iglehart, J.D. p53 alterations in all stages of breast cancer. *J. Surg. Oncol.* **1991**, *48*, 260–267.

[208] Kuntz, I.D.; Blaney, J.M.; Oatley, S.J.; Langridge, R.; Ferrin, T.E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.

[209] Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410.

[210] Skehan, P.; Storeng, R.; Scudiero, D.; Monks, A.; McMahon, J.; Vistica, D.; Warren, J.T.; Bokesch, H.; Kenney, S.; Boyd, M.R. New colorimetric cytotoxicity assay for anticancer-drug screening. *J. Natl. Cancer Inst.* **1990**, *82*, 1107–1112.

[211] Liang, Y.; Hyder, S.M. Proliferation of endothelial and tumor epithelial cells by progestin-induced vascular endothelial growth factor from human breast cancer cells: paracrine and autocrine effects. *Endocrinology* **2005**, *146*, 3632–3641.

[212] Liang, Y.; Brekken, R.A.; Hyder, S.M. Vascular endothelial growth factor induces proliferation of breast cancer cells and inhibits the anti-proliferative activity of anti-hormones. *Endocr. Relat. Cancer* **2006**, *13*, 905–919.

[213] Thoma, R.; Schulz-Gasch, T.; D'Arcy, B.; Benz, J.; Aebi, J.; Dehmlow, H.; Hennig, M.; Stihle, M.; Ruf, A. Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. *Nature* **2004**, *432*, 118–122.

[214] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

[215] Morand, O.H.; Aebi, J.D.; Dehmlow, H.; Ji, Y.H.; Gains, N.; Lengsfeld, H.; Himber, J. Ro 48-8.071, a new 2,3-oxidosqualene:lanosterol cyclase inhibitor lowering plasma

cholesterol in hamsters, squirrel monkeys, and minipigs: comparison to simvastatin. *J. Lipid Res.* **1997**, *38*, 373–390.

VITA

Sam Grinter was born in Missouri. He received Bachelors of Science in Mathematics and Physics from the University of Missouri. He continued his studies in the PhD program of the University of Missouri Informatics Institute. During his doctoral education, he researched protein–ligand docking methodology and application, as well as protein structure prediction. He received a PhD in Informatics in May of 2014.