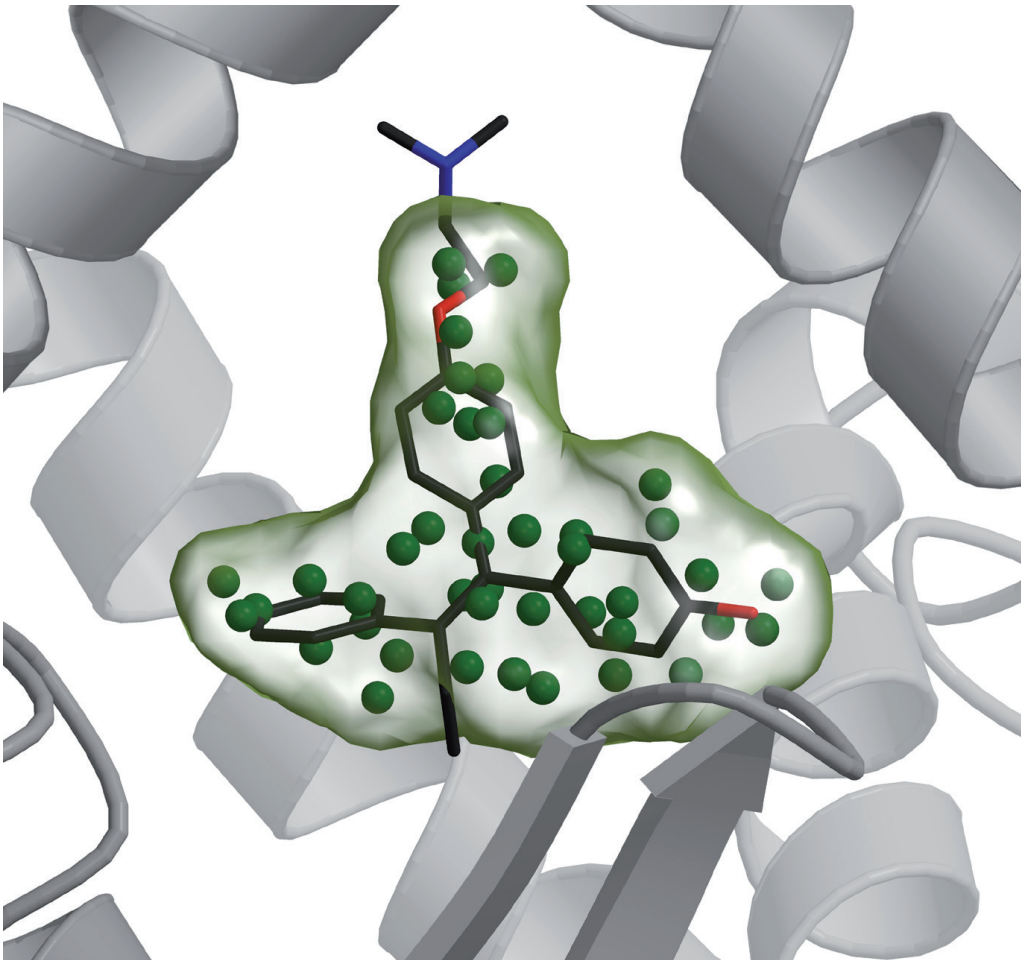
JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE
308

Sanna Niinivehmas

Computational Studies of Biomolecular Screening and Interactions



JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE 308

Sanna Niinivehmas

Computational Studies of Biomolecular Screening and Interactions

Esitetään Jyväskylän yliopiston matemaattis-luonnontieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Ylistönrinteellä salissa YAA303
marraskuun 27. päivänä 2015 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Mathematics and Science of the University of Jyväskylä,
in Ylistönrinne, hall YAA303, on November 27, 2015 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2015

Computational Studies of Biomolecular Screening and Interactions

JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE 308

Sanna Niinivehmas

Computational Studies of Biomolecular Screening and Interactions

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2015

Editors
Varpu Marjomäki
Department of Biological and Environmental Science, University of Jyväskylä
Pekka Olsbo, Sini Tuikka
Publishing Unit, University Library of Jyväskylä

Jyväskylä Studies in Biological and Environmental Science
Editorial Board

Jari Haimi, Anssi Lensu, Timo Marjomäki, Varpu Marjomäki
Department of Biological and Environmental Science, University of Jyväskylä

Cover picture by Sanna Niinivehmas

URN:ISBN:978-951-39-6382-8
ISBN 978-951-39-6382-8 (PDF)

ISBN 978-951-39-6381-1 (nid.)
ISSN 1456-9701

Jyväskylä University Printing House, Jyväskylä 2015

"You've picked up a bit of an attitude, still
curious and willing to learn, I hope."

Cheshire Cat to Alice
In Alice in Wonderland

ABSTRACT

Niinivehmas, Sanna
Computational studies of biomolecular screening and interactions
Jyväskylä: University of Jyväskylä, 2015, 79 p.
(Jyväskylä Studies in Biological and Environmental Science
ISSN 1456-9701; 308)
ISBN 978-951-39-6381-1 (nid.)
ISBN 978-951-39-6382-8 (PDF)
Yhteenveto: Laskennallisia tutkimuksia biomolekyylien seulonnasta ja vuorovaikutuksista
Diss.

Computational methods have a crucial role in modern drug discovery and design. Virtual screening (VS) is an efficient and cost-effective way to screen large molecular databases computationally in the search for novel drug candidate molecules. In VS, the commonly used computational methods can be categorized into ligand-based and protein structure-based methods. Ligand-based methods rely on the idea that similar compounds have similar properties, while protein structure-based methods utilize information about the target protein when ligand binding is predicted. This doctoral thesis continues the development of Panther, a promising novel negative image-based VS method. This VS method employs chemical and structural information derived from the ligand-binding area of the protein in the very rapid screening of molecules. In addition, several other computational methods are tested for their ability to be used as VS tools and to predict ligand binding. This comparison of the VS methods showed differences in their ability to identify active molecules and predict binding affinities, which helped to determine the relative advantages and limitations of these methods. However, based on these results, it is difficult to recommend one method over other methods.

Keywords: Binding free energy; computational drug discovery; molecular docking; molecular dynamics; negative image-based screening; virtual screening.

Sanna Niinivehmas, University of Jyväskylä, Department of Biological and Environmental Science/Nanoscience Center, P.O. Box 35, FI-40014 University of Jyväskylä, Finland

Author's address

Sanna Niinivehmas
Department of Biological and Environmental Science/
Nanoscience Center
P.O. Box 35
FI-40014 University of Jyväskylä
Finland
sanna.niinivehmas@jyu.fi

Supervisor

Olli Pentikäinen, Ph.D., Adjunct Professor
Department of Biological and Environmental Science/
Nanoscience Center
P.O. Box 35
FI-40014 University of Jyväskylä
Finland

Reviewers

Tuomo Laitinen, Ph.D., Adjunct Professor
University of Eastern Finland
School of Pharmacy
P.O. Box 1627
FI-70211 Kuopio
Finland

Adyary Fallarero, Ph.D., Adjunct Professor
Faculty of Pharmacy
P.O. Box 56
FI-00014 University of Helsinki
Finland

Opponent

Daniela Schuster, Priv.-Doz., Dr.
Institute of Pharmacy/Pharmaceutical Chemistry
Computer-Aided Molecular Design Group
University of Innsbruck
Innrain 80-82
6020 Innsbruck
Austria

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

The thesis is based on the following original articles which are referred to in the text by their Roman numerals.

I Niinivehmas S.P.*, Virtanen S.I.*, Lehtonen J.V., Postila P.A. & Pentikäinen O.T. 2011. Comparison of virtual high-throughput screening methods for the identification of phosphodiesterase-5 inhibitors. *Journal of Chemical Information and Modeling* 51: 1353-1363.

II Niinivehmas S.P.*, Manivannan E.*, Rauhamäki S., Huuskonen J. & Pentikäinen O.T. 2015. Identification of estrogen receptor α ligands with virtual screening techniques. Submitted manuscript.

III Niinivehmas S.P., Salokas K., Lätti S., Raunio H. & Pentikäinen O.T. 2015. Ultrafast protein structure-based virtual screening with Panther. *Journal of Computer-Aided Molecular Design* 10: 989-1006.

IV Virtanen S.*, Niinivehmas S.P.* & Pentikäinen O.T. 2015. Case-specific performance of MM-PBSA, MM-GBSA, and SIE in virtual screening. *Journal of Molecular Graphics and Modelling*. In press. doi:10.1016/j.jmgm.2015.10.012.

V Shubina V., Niinivehmas S.P. & Pentikäinen O.T. 2015. Reliability of virtual screening methods in prediction of PDE4B-inhibitor activity. *Current Drug Discovery Technologies* 12: 117-26.

VI Sergelius C., Niinivehmas S., Maula T., Kurita M., Yamaguchi S., Yamamoto T., Katsumura S., Pentikäinen O.T. & Slotte J.P. 2012. Structure-activity relationship of sphingomyelin analogs with sphingomyelinase from *Bacillus cereus*. *Biochimica et Biophysica Acta* 1818: 474-480.

* equal contribution.

RESPONSIBILITIES OF SANNA NIINIVEHMAS IN THE THESIS ARTICLES

Article I — I performed the studies for phosphodiesterase type 5. Salla Virtanen performed the studies for other targets. I prepared the figures together with Pekka Postila. I wrote the article together with other authors. This article has been previously used in the thesis of Salla Virtanen.

Article II — I made all negative image-based studies and docking with PLANTS. Elangovan Manivannan made other computational studies. Sanna Rauhamäki performed fluorescence polarization measurements. Juhani Huuskonen synthesized the molecules. I wrote the manuscript with other authors and prepared most of the figures.

Article III — I conducted the virtual screening studies, prepared the figures and wrote the manuscript apart from the model creation part, which was written by Kari Salokas.

Article IV — I implemented the virtual screening studies and binding free energy calculations for β-lactamase. Other computational studies were performed by Salla Virtanen. I revised the manuscript together with other authors. The earlier version of the manuscript was used in the thesis of Salla Virtanen.

Article V — I completed the computational studies together with Victoria Shubina. I wrote the article together with other authors.

Article VI — I performed the computational studies and prepared the figures and wrote the sections related to these experiments.

ABBREVIATIONS

1D	one dimensional
2D	two dimensional
3D	three dimensional
3D-QSAR	3D-quantitative structure-activity relationship
ADMET	Absorption Distribution Metabolism Excretion Toxicity
ALR2	aldose reductase 2
AmpC	β-lactamase
AUC	area under curve
DS	dataset
DUD	Directory of Useful Decoys
DUD-E	Database of Useful Decoys: Enhanced
ERα	estrogen receptor alpha
HSP90	human heat shock protein 90
MD	molecular dynamics
MMGBSA	molecular mechanics generalized Born surface area
MMPBSA	molecular mechanics Poisson-Boltzmann surface area
NIB	negative image-based
NMR	nuclear magnetic resonance
PAINS	pan assay interference compounds
PDB	Protein Data Bank
PDE5	cyclic nucleotide phosphodiesterase type 5
PR	progesterone receptor
RMSD	root mean square deviation
ROC	receiver operating characteristics
SERM	selective estrogen receptor modulator
SIE	solvated interaction energy
SM	sphingomyelin
SMase	sphingomyelinase
VS	virtual screening

1 INTRODUCTION

Prescribed in the diagnosis, treatment, cure, and prevention of diseases, drugs are chemical substances that cause therapeutic effects in patients. The most commonly active substances in drugs are small organic molecules that act by modulating the function of the proteins linked to the metabolic or signaling pathways related to diseases. Proteins are large and diverse biological macromolecules that are essential in virtually every process occurring within cells. For example, enzymes catalyze chemical reactions, and receptor proteins convey signals to and from the cell as well as between the cytoplasm and the nucleus within the cell. Because of the versatility of proteins, drugs can affect them in many different ways. Drugs usually either mimic the action of the natural ligand, thereby enhancing the function of the protein (agonist or substrate), or they block the function of the natural ligand, thereby preventing or reducing the function of the protein (antagonist or inhibitor).

Launching a new drug in the market is a long and complex process. Pharmaceutical research projects typically consist of two parts: discovery and development. In the discovery phase, hundreds of thousands of molecules can be studied by using a high-throughput screening approach to reveal their biological activity in a selected target protein panel. However, only a few of the tested compounds are promising enough to enter the development phase in which the compounds are assessed and optimized for their efficacy and clinical safety. Despite the huge amounts of effort, time, and money invested in pharmaceutical research, only about 10% to 15% of the molecules entering clinical trials become accepted as drugs. This fact supports the view that rational drug discovery requires methods that both speed up and increase the effectiveness of the discovery process used to identify potent compounds.

Virtual screening (VS) is a cost-effective, fast approach to screening huge molecular databases computationally in the search for novel drug candidate molecules. VS enables experimental testing to focus on the top hit molecules, thus offering a way to enhance drug discovery. VS methods can be designed to meet specific requirements, depending on the available data, such as the known ligands and/or information about the structure of target proteins. VS can be

effectively applied not only to the discovery of novel hit and lead compounds but also to the prediction of properties of compounds in metabolism and toxicity, for example. This doctoral thesis continues the development of the promising novel negative image-based VS method Panther. This method utilizes structural and chemical information about the ligand-binding area of the protein in the very rapid screening of molecules. In addition, several other computational methods are tested for their ability to be used as VS tools and predict ligand binding.

2 REVIEW OF LITERATURE

2.1 Virtual screening

In virtual screening (VS), the aim is to identify bioactive compounds in large molecular databases through computational means and to enrich the most potent molecules in the top fraction for experimental testing. Because VS is a powerful strategy for identifying hit molecules as starting points in medicinal chemistry, it has become a standard procedure in drug discovery. For that reason, the development of VS approaches has gained attention, and several comparative studies of the usability of various methods have been published (Sheridan and Kearsley 2002, Warren *et al.* 2006, McGaughey *et al.* 2007, Kolb and Irwin 2009, von Korff *et al.* 2009, Krüger and Evers 2010, Venkatraman *et al.* 2010, Plewczynski *et al.* 2011, Scior *et al.* 2012, Heikamp and Bajorath 2013, Grinter and Zou 2014, Danishuddin and Khan 2015). VS methods comprise of several computational techniques that, depending on the utilized information, can be classified into two broad categories: ligand-based and protein structure-based.

Different VS approaches are able to retrieve potent compounds with different characteristics (Sheridan and Kearsley 2002, Zhang and Muegge 2006, Chen *et al.* 2009, von Korff *et al.* 2009, Krüger and Evers 2010, Venkatraman *et al.* 2010). Choosing the best-suited method for the VS study depends on many factors, such as the availability of high-quality structural data on the target protein, existing drugs or known active ligands, accessible computational resources, and the size of the database to be screened. However, the question concerns not only the method selection but also the tools that can or should be used serially or in parallel to screen databases more efficiently (Tan *et al.* 2008, Krüger and Evers 2010, Svensson *et al.* 2012, Drwal and Griffith 2013). The common practice is to utilize different VS methods serially so that first computationally more efficient methods are used to screen out the greater part of the database molecules; and then the remaining top-ranked compounds are screened using more accurate and computationally more strenuous methods (Fig. 1).

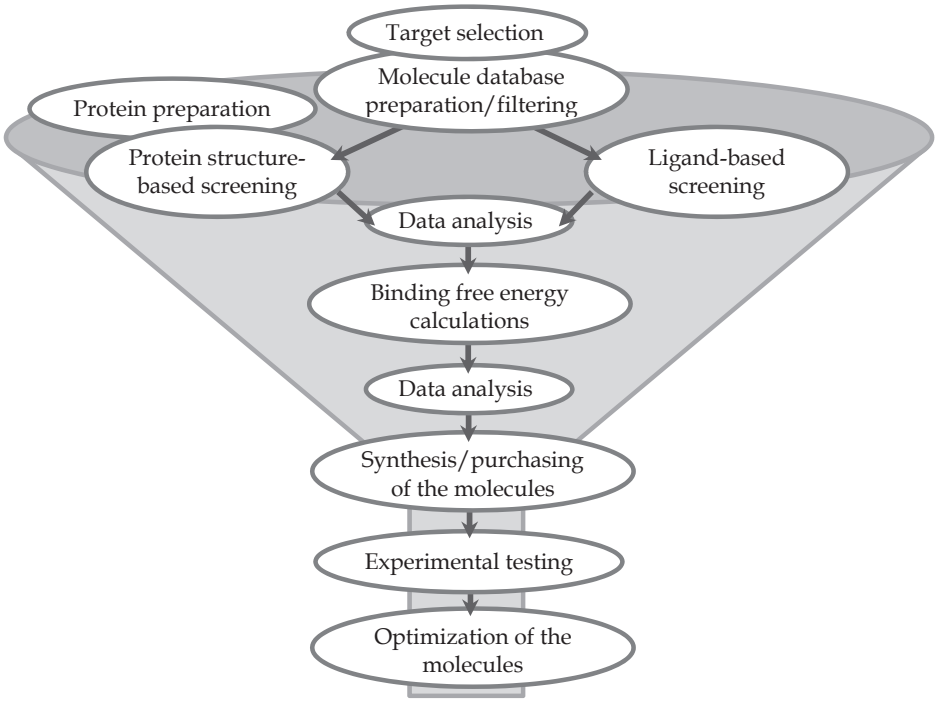


FIGURE 1 Overview of the virtual screening procedure. Typically, the study proceeds from the utilization of more approximate methods to the utilization of more accurate methods. Ligand-based and protein structure-based methods also can be used simultaneously.

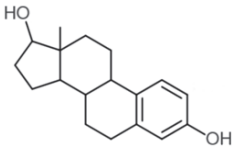
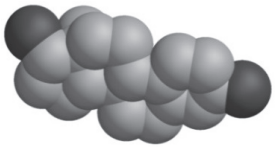
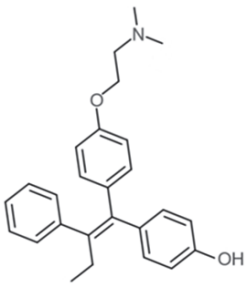
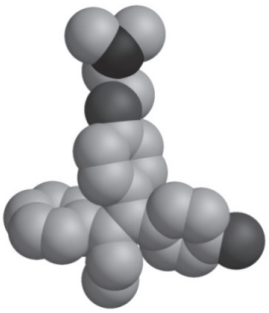
Several studies have reported the successful utilization of VS methods to determine lead molecules (Ghosh *et al.* 2006, Cavasotto *et al.* 2008, Clark 2008, Markt *et al.* 2008, Ripphausen *et al.* 2010, Murgueitio *et al.* 2012, Kumar and Zhang 2014, Cerqueira *et al.* 2015, Noha *et al.* 2015). However, the methods still require improvement in accuracy and computational efficiency (Warren *et al.* 2006, Clark 2008, Cheng *et al.* 2012, Scior *et al.* 2012, Heikamp and Bajorath 2013, Lavecchia and Di Giovanni 2013, Zhu *et al.* 2013, Cerqueira *et al.* 2015, Chen 2015). The principles, strengths, and weaknesses of ligand-based methods, molecular docking, pharmacophore modeling, structure-activity relationship models, negative image methods, and binding free energy calculation methods are discussed in the following sections.

2.1.1 Ligand-based methods

Ligand-based methods rely on the idea that similar compounds have similar properties, which may lead to similar biological functionality (Johnson *et al.* 1989, Patterson *et al.* 1996). In ligand-based methods, small molecules are characterized by molecular descriptors or fingerprints, which are distinctive characteristics or unique patterns indicating the properties of the molecules. Different molecular descriptors can be categorized for example according to the

dimensionality (Table 1). Simple one-dimensional (1D) descriptors typically consist of bulky properties such as molecular weight, octanol/water partition coefficient, or the number of particular atoms. However, 1D line notations, such as SMILES, may contain also structural information. Two-dimensional (2D) descriptors utilize the chemical structure of which e.g. substructures, the number of bonds or connectivity information can be derived. Three-dimensional (3D) descriptors, such as shape, volume, solvent accessible surface area, and 3D pharmacophore properties, depend on conformers of molecules. 1D descriptors are used rather to filter molecular databases than screening of actual hit molecules, whereas 2D and 3D descriptors are widely used in similarity searches in VS (Zhang and Muegge 2006, McGaughey *et al.* 2007, Krüger and Evers 2010, Venkatraman *et al.* 2010, Hu *et al.* 2012, Dobi *et al.* 2014).

TABLE 1 Molecular descriptors used in similarity searches can be classified by the dimensionality. Here a natural ligand 17β-estradiol and an antagonist molecule 4-hydroxytamoxifen of estrogen receptor α are shown as examples.

Molecule and 1D structures	2D structure	3D structure
17β-estradiol $C_{18}H_{24}O_2$ C[C@]12CC[C@H]3[C@@H](CCc4cc(O)ccc34)[C@@H]1CC[C@@H]2O		
4-hydroxytamoxifen $C_{26}H_{29}NO_2$ CC\C(=C(/c1ccc(O)cc1)\c2ccc(OCCN(C)C)cc2)\c3ccccc3		

The most commonly used structural 2D fingerprints (or topological descriptors) indicate for example either the presence or the absence of particular substructures in the molecule. The most common measure to compare similarities of these fingerprints is to calculate the Tanimoto coefficient (Bajusz *et al.* 2015). The benefit of methods that use 2D fingerprints in VS is their computational efficiency due to no specific information about conformers is needed in the screening process (Sheridan and Kearsley 2002, Venkatraman *et al.* 2010, Hu *et al.* 2012). Another common approach in ligand-based VS is 3D screening based on the shape of the molecules. In addition to using the shape of the molecules, chemical features, such as pharmacophoric properties or electrostatic potentials, can be taken into account. Some commonly used 2D and 3D software in ligand-based similarity searching are listed in Table 2.

TABLE 2 Commonly used software in ligand-based virtual screening.

Software	Application area	References
BCI	2D fingerprints	Barnard and Downs 1997
Canvas	2D fingerprints	Duan *et al*. 2010, Sastry *et al*. 2010
ChemAxon Screen	2D fingerprints, 3D shape, pharmacophore	ChemAxon Kft., Budabest, Hungary
Daylight	2D fingerprints	Daylight Chemical Information Systems Inc., Laguna Niguel, CA
Discovery Studio	2D fingerprints, 3D similarity, pharmacophore, 3D-QSAR	Accelrys Inc., San Diego, CA
ECFP	2D fingerprints for structure-activity modeling	Rogers and Hahn 2010
ElectroShape	3D shape and electrostatics	Armstrong *et al*. 2010, Armstrong *et al*. 2011
LigandScout	pharmacophore	Wolber and Langer 2005, Wolber *et al*. 2006
MACCS	2D fingerprints	MDL Information Systems Inc., San Leandro, CA
MOE	2D fingerprints, 3D similarity, pharmacophore, 3D-QSAR	Chemical Computing Group, Montreal, Canada
Molprint2D	2D fingerprints	Bender *et al*. 2004
Open Babel	2D fingerprints	O'Boyle *et al*. 2011
Phase	3D similarity, pharmacophore, 3D-QSAR	Dixon *et al*. 2006a, Dixon *et al*. 2006b
ROCS	3D shape, pharmacophore	Rush *et al*. 2005
ShaEP	3D shape and electrostatics	Vainio *et al*. 2009
SiFT	3D protein-ligand interactions as 1D fingerprints	Deng *et al*. 2004
Unity	2D fingerprints, 3D shape, pharmacophore	Tripos Inc., St. Louis, MO
USR	3D shape	Ballester and Richards 2007, Ballester *et al*. 2009, Ballester 2011

In general, ligand-based methods require relatively little computational capacity, memory, and data storage. Compared to 2D methods, computationally 3D methods are substantially more strenuous because of their increased complexity and the demand for handling multiple low-energy conformers; however, the processing time per molecule is still reasonable for VS (Sheridan and Kearsley 2002, Zhang and Muegge 2006, McGaughey *et al*. 2007, Cortés-Cabrera *et al.* 2013). The benefit of 3D methods over topological searches is that they offer more information about the possible bioactive binding conformer, which is crucial in the activity of a compound (Sheridan and Kearsley 2002, Zhang and Muegge 2006, Rogers and Hahn 2010).

Because ligand-based methods depend heavily on the reference structure (or multiple queries), the selection of the query greatly influences the success of VS (Kirchmair *et al.* 2009). Exploiting different reference ligands may drastically change the effectiveness of ligand-based VS, and thus, the effectiveness of

methods used to ligand-based similarity searches can vary in ways that are difficult to predict (Sheridan and Kearsley 2002, Ripphausen *et al.* 2011). The application of multiple diverse ligands generally leads to higher scaffold hopping and improves the enrichment of potential molecules because dissimilar queries tend to select different subsets of active molecules (although often analogous to the reference structures) from databases (Sheridan and Kearsley 2002, Krüger and Evers 2010, Ripphausen *et al.* 2011). Utilizing only the shape of the molecules variable scaffolds are found efficiently but effective interactions with the target protein are neglected. However, this can be overcome with the optimization of functional groups of the molecules later. On the other hand, topological methods do find active properties, but found molecules are not necessarily best fitted to the ligand-binding area of the target protein. In any case, the comparison of molecular similarities is one of the simplest and most widely used methods in VS, and it has been shown that ligand-based screening methods are efficient in finding active molecules (Evers *et al.* 2005, Markt *et al.* 2008, Mochalkin *et al.* 2009, Ripphausen *et al.* 2011, Simões *et al.* 2010, Du-Cuny *et al.* 2011, Hevener *et al.* 2012, Dobi *et al.* 2014).

2.1.2 Protein structure-based methods

Protein structure-based methods (Fig. 2), of which the most common is molecular docking, have been widely used ever since the early 1980s (Kuntz *et al.* 1982). Consequently, the number of available docking algorithms is very large (see Table 5 below). In molecular docking, the binding mode of a ligand within the binding site of a protein is predicted, and an estimation of the binding affinity is obtained. This is done by sampling several poses of small molecules positioned in the ligand-binding area of the 3D representation of the protein structure and then assessing the complementarity of their steric and chemical features. Typically, although in docking ligands are treated flexibly and the protein structure is fixed, it is also possible to use rigid ligand conformers or to let protein structure adapt. In addition to the flexibility of the protein-ligand complex, it is also important to consider the water molecules that mediate essential interactions.

Protein conformation can change upon ligand binding and because of that docking is sensitive to the protein structure. When docking is performed in a fixed protein structure, it may lead into degeneration or even total failure of the docking process. The direct modeling of protein binding site flexibility is very challenging because of the complex conformational space. Another option for modeling protein adaptability is to use an ensemble of rigid protein conformers that are either determined experimentally by X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy or are generated computationally via molecular dynamics (MD) simulations. Obviously, considering protein flexibility increases the amount of computing time required to complete the docking process (Ferrara *et al.* 2004, Sheridan 2008, Totrov and Abagyan 2008, Lill 2011, Therrien *et al.* 2014).

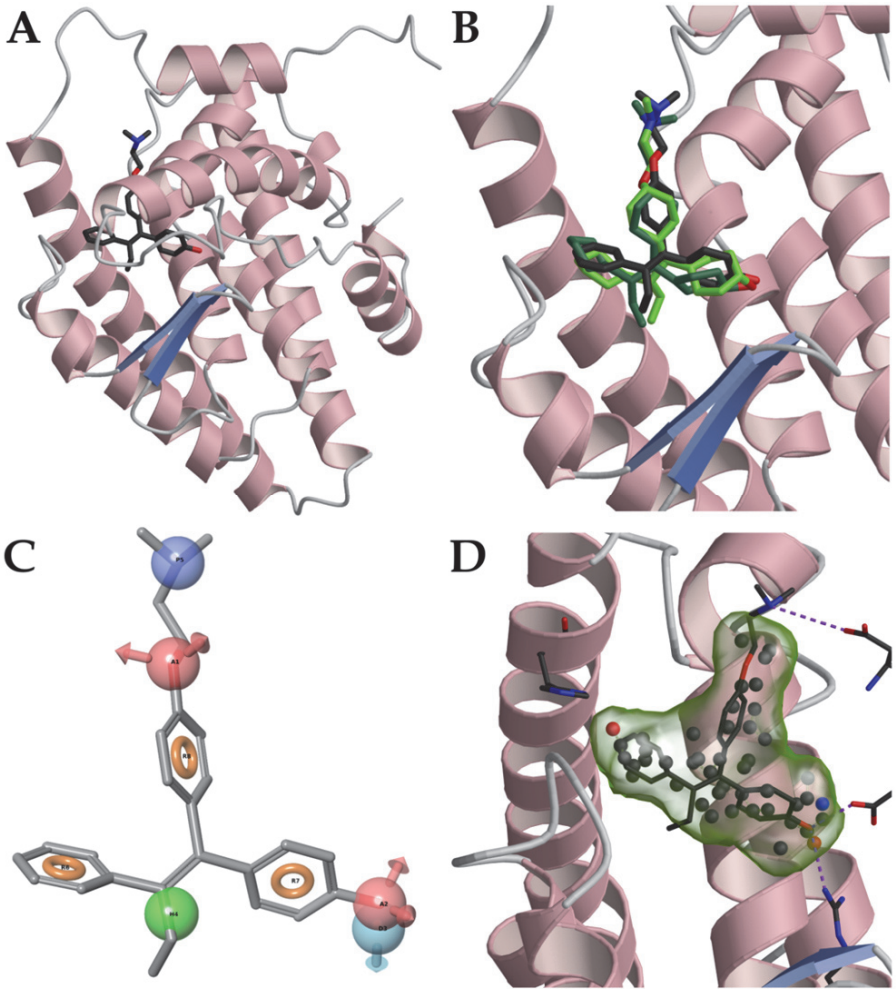


FIGURE 2 Protein structure-based virtual screening methods. A) High quality protein structures work as a starting point in protein structure-based virtual screening studies. Here the estrogen receptor α (ERα) is represented in a secondary structure as an example of an X-ray crystal structure (PDB: 3ert, Shiau *et al*. 1998). B) Molecular docking, molecular dynamics simulations, and binding free energy calculation methods can predict small molecule binding poses. Here is a close-up of the ERα ligand-binding area with 4-hydroxytamoxifen from X-ray structure (black carbon atoms), docked (dark green carbon atoms), and after binding free energy calculations (light green carbon atoms). C) A pharmacophore model of ERα with 4-hydroxytamoxifen as a reference ligand. The red and blue arrows indicate hydrogen bond acceptors and donors, respectively. The golden toroids designate aromatic rings. The darker blue shows a positive group, and the green sphere points the hydrophobic area. D) In negative image methods, the shape and the properties of the ligand-binding area of the protein are reflected in the model. The green surface outlines the shape of the negative image model. The red balls represent negatively charged model points, the blue balls represent positively charged model points, and the grey balls represent neutral model points.

When a protein molecule is in solution, the surface and the curvatures are entirely embedded in water molecules. Although most of the water molecules are mobile and only loosely bound to protein, some water molecules are more important than others are, and they mediate the interactions between the ligand and the protein by forming hydrogen-bond networks that stabilize the complex. The importance of water molecules depends on the binding ligand and thus the significance of the water molecules needs to be defined on a case-by-case basis (Roberts and Mancera 2008, Thilagavathi and Mancera 2010, Lie *et al.* 2011, Therrien *et al.* 2014). Other factors to consider include the modeling of interactions with a possible metal or cofactor at the ligand-binding area and that both the ligand-binding mode and affinity can be strongly pH-dependent (Ferrara *et al.* 2004).

Different ligand conformers can be generated for molecular docking, either by using an external conformer generation method prior to the docking procedure or, more commonly, a method that is implemented in a docking program. Molecular docking programs (as well as other conformer generation methods) use different search algorithms to explore the conformational space for flexible ligand docking. The methods used to treat ligand flexibility can be divided into three basic categories: systematic, stochastic, and simulation-based (Kitchen *et al.* 2004, Huang and Zou 2010, Table 3).

TABLE 3 Conformational search methods.

Category and method	Working principle
Systematic - Incremental construction	Degrees of freedom of a molecule are explored exhaustively i.e. the conformational search space consists of all possible conformers and orientations. Typically ligand fragments or core part (rigid) are docked and then torsion angles of flexible parts are systematically sampled before linking the docked parts.
Stochastic/random - Genetic algorithm - Monte Carlo simulation	Random changes are made to a ligand or a ligand tree i.e. a random starting conformer is generated and evaluated, and then a second randomly generated conformer is compared with the first one and the more optimal conformer is chosen for continuation. By repeating these cycles, the conformer of the molecule becomes optimized.
Simulation-based - Molecular dynamics - Energy minimization	A molecule is allowed to freely explore the conformational space in the ligand-binding area, and the force on each atom is calculated by the change in the potential energy between the current and new positions of the atoms.

Both the posing and the ranking of predicted ligand conformers are vital aspects of molecular docking. Scoring functions are used to evaluate the fitness of the suggested ligand-binding poses and estimate the binding free energy in protein-ligand interactions. Scoring functions are mathematical approximation

methods that are based on several simplifications and assumptions that are made when the protein-ligand complexes are assessed. Scoring functions need to distinguish active molecules and conformers from inactive ones, rank poses correctly, and arrange them according to their predicted binding affinity. There are three types of scoring functions: force field-based, empirical, and knowledge-based (Halperin *et al*. 2002, Ferrara *et al*. 2004, Kitchen *et al*. 2004, Huang and Zou 2010; Table 4). In addition to aforementioned scoring function types, also approach called consensus scoring is utilized. In consensus scoring, several scoring functions or their components are combined, and only the poses that fulfill all the required conditions are accepted.

TABLE 4 Scoring functions.

Type	Working principle
Force field-based	Molecular mechanics energy functions are applied. Binding free energy is approximated as a sum of intermolecular van der Waals interactions and electrostatic interactions. Also the effects of solvation and intramolecular energies can be included.
Empirical	Experimental binding constants of a training set of protein-ligand complexes are utilized. Binding free energy is estimated as a sum of interaction terms that consist of contributions of hydrogen bonds, ionic interactions, hydrophobic effect, and entropy.
Knowledge-based	Binding affinity is a sum of protein-ligand atom pair interactions derived from statistical analysis of known complexes. Probability of atoms or functional groups to lie close to each other is converted to distance-dependent interaction score. Frequently occurring interactions are assumed to be energetically more favorable.

One reason for the popularity of docking is that the number of high-quality protein structures has increased dramatically in recent years (Berman *et al.* 2000, Rose *et al.* 2015). Because of the complexity and high dimensionality of the conformational space, the docking approach is generally a computationally demanding procedure and requires significantly more computing time than ligand-based methods (Krüger and Evers 2010). However, the steady increase in computing power has enabled these approaches to become standard computational tools in structure-based VS. Although molecular docking is widely used, not only in VS but also in the structure-based optimization of lead compounds, it still is not fully trouble-free (Grinter and Zou 2014). Some commonly used docking software and descriptions of their conformational search methods and scoring functions are listed in Table 5.

TABLE 5 Commonly used software in structure-based virtual screening.

Software	Conformational search method	Scoring function	References
AutoDock	Genetic algorithm	Empirical	Goodsell *et al*. 1996, Morris *et al*. 2009
DOCK	Incremental construction	Force field-based	Kuntz *et al*. 1982, Lang *et al*. 2009
eHITS	Exhaustive systematic	Empirical, knowledge-based	Zsoldos *et al*. 2007, Ravitz *et al*. 2011
FlexX	Incremental construction	Empirical	Rarey *et al*. 1996, Kramer *et al*. 1999
FRED	Exhaustive shape matching of multiple rigid conformers	Gaussian, empirical	McGann *et al*. 2003, McGann 2011
Glide	Incremental construction with Monte Carlo sampling	Combined empirical and force field-based	Friesner *et al*. 2004, Halgren *et al*. 2004, Friesner *et al*. 2006
GOLD	Genetic algorithm	Empirical, force field-based	Jones *et al*. 1997, Verdonk *et al*. 2003
LigandFit	Shape complementarity with Monte Carlo sampling	Empirical	Venkatachalam *et al*. 2003, Krammer *et al*. 2005
PLANTS	Ant colony optimization	Empirical	Korb *et al*. 2009
Surflex	Incremental construction and surface-based similarity	Empirical	Jain 2003, Cleves and Jain 2015

Docking programs are usually rather successful in identifying ligand poses that resemble crystallographically determined or bioactive ligand conformers; however, problems sometimes occur in the creation of the bioactive ligand conformers (Warren *et al.* 2006, Kolb and Irwin 2009). Due to the inability of the current scoring functions to distinguish and differentiate sometimes subtle differences that can change ligand affinity from highly potent to inactive docking does not always work for the right reasons (Warren *et al.* 2006). However, a more wide-ranging issue is that docking programs are not accurate and reliable enough when predicting binding affinities (Ferrara *et al.* 2004, Warren *et al.* 2006, McGaughey *et al.* 2007, Cross *et al.* 2009, von Korff *et al.* 2009, Feliu and Oliva 2010, Plewczynski *et al.* 2011).

An additional concern is that docking results are quite often case specific, and no single program performs well for all targets (Ferrara *et al.* 2004, Warren *et al.* 2006, McGaughey *et al.* 2007, Cross *et al.* 2009, von Korff *et al.* 2009, Feliu and Oliva 2010, Plewczynski *et al.* 2011). This shortcoming may require a considerable contribution by the researcher. In particular, extensive validation is required if no prior knowledge is available about the suitability of the program for certain targets. In general, if the applicability of the docking program is evaluated carefully, and it is proved that it is capable of separating active compounds from a pool of inactive compounds, its usage as a VS tool is supported (Ferrara *et al.* 2004, Warren *et al.* 2006, McGaughey *et al.* 2007, Cross *et al.* 2009, von Korff *et al.* 2009, Feliu and Oliva 2010, Plewczynski *et al.* 2011).

2.1.3 Pharmacophore modeling

Popular VS methods utilizing 3D descriptors include pharmacophore models (Table 2; Fig. 2C). Earliest modern pharmacophore techniques date back to 1970s (Martin *et al.* 1973, Gund 1979), and since then many applications of pharmacophores have been developed. Nevertheless, the basic concept of a pharmacophore as a simple geometric representation of the key molecular interactions has remained unchanged. For the evolution of the pharmacophore concept see (Güner 2002, Leach *et al.* 2010, Yang 2010).

Pharmacophore models represent molecular features that are known or thought to be crucial in protein-ligand complex formation (Fig. 2C). Pharmacophore models try to find common features and thus explain how structurally diverse ligands can bind to the same ligand-binding area of the protein (Güner 2002, Leach *et al.* 2010, Yang 2010). Characteristic pharmacophoric features consist of steric and electrochemical properties, such as aromatics rings, hydrophobic interactions, and hydrogen bond acceptors and donors (Fig. 2C). Pharmacophore models can be derived either by using an ensemble of known ligand molecules or by projecting from the protein structure (Güner 2002, Leach *et al.* 2010, Yang 2010). For the creation of a ligand-based pharmacophore model, a set of structurally diverse compounds is selected as training molecules. The training molecules need to be superimposed, and they should contain a bioactive conformer. Based on the overlay of the training molecules, the common pharmacophoric points are identified. The validity of the model can be tested by screening an ensemble of test molecules whose activity status is known to determine whether the pharmacophore model can predict them correctly. In general, the pharmacophore model identifies and separates the active and inactive molecules by extrapolating the knowledge acquired from the training set compounds (Dixon *et al*. 2006a, Dixon *et al*. 2006b, Yang 2010). The molecular database can then be searched for more molecules that share the same features arranged in the same relative orientation. Although the generation of a pharmacophore model can be time-consuming and require diligence, VS with pharmacophore models is fast and efficient (Güner 2002, Schuster *et al.* 2008, Chen *et al.* 2009, Leach *et al.* 2010, Yang 2010, Vuorinen *et al.* 2015).

2.1.4 3D-quantitative structure-activity relationship

In 3D-quantitative structure-activity relationship (3D-QSAR) relationship between the biological activity of a set of compounds and their 3D properties are determined quantitatively using statistical correlation methods (Hansch and Fujita 1964, Hansch *et al.* 1965, Hansch 1969). The basis for the technique lies on a simple assumption that similar molecules have similar biological activity. The first widely used 3D-QSAR method, comparative molecular field analysis (CoMFA), was introduced in 1988 (Cramer *et al.* 1988). For the advancement of the QSAR concept see (Cherkasov *et al.* 2014).

In 3D-QSAR is regression models are used to study the correlation between chemical structure and the biological activity of small molecules (Kubinyi 1997, Verma *et al.* 2010). For the creation of the 3D-QSAR model, comparable data with a wide range of affinity divided into training and test set are needed (Golbraikh *et al.* 2003). In principle, the same phases as in pharmacophore models apply (see above), and pharmacophore models often can be used as a starting point in developing 3D-QSAR models (Dixon *et al.* 2006a, Dixon *et al.* 2006b). The difference is that 3D-QSAR models utilize the knowledge acquired from known molecules to predict the activity values of new compounds. The 3D-QSAR is not only used for drug discovery, lead optimization, and binding affinity predictions but also is applied to toxicity predictions, risk assessment, and regulatory decisions. The disadvantage of both the pharmacophore model and the 3D-QSAR model is that they depend strongly on the quality of the input data, and hence derived hypotheses usually rely on a finite number of chemical data. Furthermore, the models cannot predict activities that are outside the range of the training dataset, or cannot identify novel chemical structures that are not already presented during the model creation (Kubinyi 1997, Golbraikh *et al.* 2003, Verma *et al.* 2010).

2.1.5 Negative image of the ligand-binding area

In negative image-based (NIB) screening, the ligand-binding area of the target protein is represented as a ligand-like entity, which is then used instead of known ligand molecule(s) in similarity searches (Fig. 2D). The idea of searching ligand-sized cavities for the purpose of drug design is not new, and it is on display from time to time (see e.g. Kleywegt and Jones 1994, Liang *et al.* 1998, Nayal and Honig 2006, Coleman and Sharp 2010, Hetényi and van der Spoel 2011). Moreover, previous studies have introduced methods that utilize negative images in VS (Oshiro and Kuntz 1998, Fukunishi *et al.* 2006, Ebalunode *et al.* 2008, Lee *et al.* 2009, Virtanen and Pentikäinen 2010, Lee and Zhang 2012). Common to all these methods is that a negative image is created of the ligand-binding area, and the model is then used to identify potential novel compounds.

Ligand-based screening is fast, but the ability of the protein to adapt according to the binding ligand is neglected. In ligand-based VS, multiple known ligands may be used to represent the adaptability of the protein; however, the results are restricted to scaffolds similar to reference ligands. Molecular docking is by far the most common method used in protein structure-based VS. However, compared to NIB screening, which also utilizes protein structure, docking is slow and computationally intensive. It has been shown that the parallel utilization of different ligand- and protein structure-based VS methods is advantageous and promotes the identification of a greater number of active and diverse compounds (Tan *et al.* 2008, Krüger and Evers 2010, Svensson *et al.* 2012, Drwal and Griffith 2013). NIB screening combines the best properties of both ligand- and protein structure-based methods by being fast and utilizing the knowledge offered by the protein needed for scaffold hopping.

This thesis presents Panther, a novel, ultrafast NIB multipurpose docking tool. In the Panther method, a simple shape-electrostatic model of the ligand-binding area of the protein is created, and then the features of possible ligands are compared to the model by using a fast similarity search algorithm (Fig. 3). Panther can be used in several applications, such as a cavity-searching tool, a VS tool in the early phases of drug discovery projects, a tool to estimate ligand-binding properties and off-target binding, as well as a tool to assess the environmental toxicity of chemicals. Panther-code is available free of charge after registration in Panther web page (http://www.jyu.fi/panther).

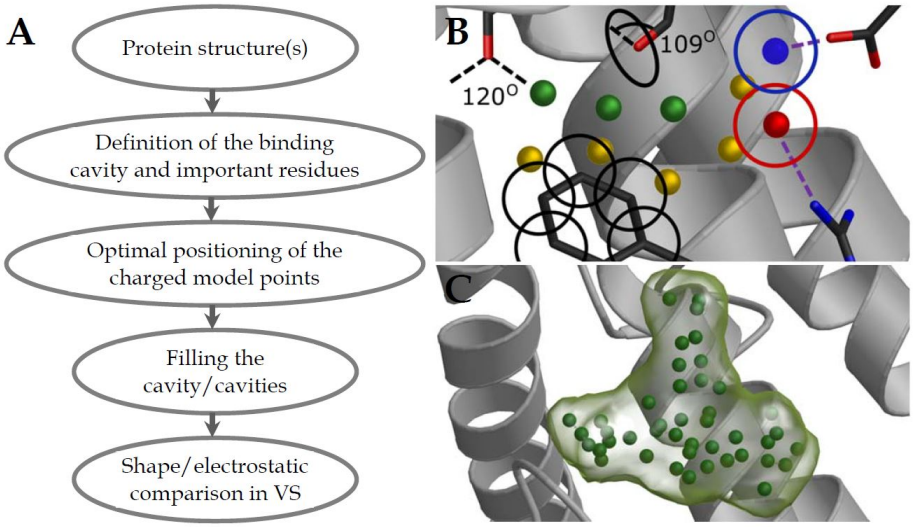


FIGURE 3 Panther modeling. A) Simplified scheme of the Panther procedure. B) Charged model atoms (red, blue) are placed around polar/charged atoms in the ligand-binding cavity with specified angle, geometry, and distance. Cavity is filled with neutral atoms (green, yellow). Overlapping atoms (yellow) are removed. C) The final model represents the negative image of the binding site.

2.2 Binding free energy calculations

Methods for the accurate prediction of the binding affinity of small molecules for a protein target are urgently needed in VS. Approximated estimations of binding affinity can be routinely made by using molecular docking programs. However, because these programs employ approximated scoring functions, the correlations between the estimated binding affinities and the experimentally measured values are often poor (Ferrara *et al.* 2004, Warren *et al.* 2006, McGaughey *et al.* 2007, Cross *et al.* 2009, von Korff *et al.* 2009, Feliu and Oliva 2010, Plewczynski *et al.* 2011). More thorough computational methods for binding free energy calculations include free energy perturbation (Kollman 1993) and thermodynamic integration (Lybrand *et al.* 1986). However, because these

methods are computationally very strenuous, reasonably accurate and computationally more efficient methods combining molecular mechanics force fields and continuum solvent models have become increasingly more widely used in binding free energy calculations (Kollman *et al.* 2000, Massova and Kollman 2000, Guimarães and Cardozo 2008). Such methods include the molecular mechanics generalized Born surface area (MMGBSA), the molecular mechanics Poisson-Boltzmann surface area (MMPBSA), and solvated interaction energy (SIE) methods.

2.2.1 Molecular mechanics generalized Born/Poisson Boltzmann surface area

In the MMGBSA and MMPBSA methods, the absolute free energy of a system is estimated from molecular mechanics (MM) energy, an estimate of the electrostatic free energy from continuum solvent model, such as generalized Born (GB) or Poisson-Boltzmann (PB), and an estimate of the solvation free energy determined from the exposed surface area (SA). The estimation of the entropy is derived from normal mode analysis; however, because of its high computational cost, it is often omitted in binding free energy calculations (Genheden and Ryde 2015).

The disadvantage of MMGBSA and MMPBSA binding free energy calculations is that in some cases they are susceptible to the length of the MD simulation. Furthermore, the method chosen for the generation of the protein-ligand complexes can alter the results (Ferrari *et al.* 2007, Hou *et al.* 2011a, Hou *et al.* 2011b, Greenidge *et al.* 2014). Another issue is that there may be significant differences between the different implicit water models, which makes the performance of the method highly case-specific (Ferrari *et al.* 2007, Hou *et al.* 2011a, Hou *et al.* 2011b). Furthermore, MMGBSA and MMPBSA are intrinsically computationally laborious. In addition, the protein-ligand complexes for MMGBSA and MMPBSA are typically acquired from MD simulations which makes the combination even more strenuous and thus impractical in VS. Therefore, the ability to estimate the binding free energy reliably by using a single energy-minimized structure and implicit solvation would promote the usage of MMGBSA and MMPBSA as VS tools. Some studies have shown that correlations obtained using a single minimized protein-ligand complex are similar to those obtained by using either using multiple protein-ligand complexes or averaging multiple MD snapshots (Ferrari *et al.* 2007, Rastelli *et al.* 2010, Sgobba *et al.* 2012, Mulakala and Viswanadhan 2013).

Many studies employing MMGBSA and MMPBSA explore only relatively small number of ligands and target proteins. Nevertheless, these studies have showed that MMGBSA or MMPBSA are able to discriminate between active molecules and inactive molecules, that docking poses can be improved by reassessing them, and that reasonable correlations between calculated binding free energies and experimental affinities can be achieved (Ferrari *et al.* 2007, Guimarães and Cardozo 2008, Rastelli *et al.* 2010, Hou *et al.* 2011a, Hou *et al.* 2011b, Mulakala and Viswanadhan 2013, Ylilauri and Pentikäinen 2013, Greenidge *et al.* 2014).

2.2.2 Solvated interaction energy

Similar to the MMGBSA and MMPBSA methods, the solvated interaction energy (SIE) (Naïm *et al.* 2007) function approximates protein-ligand binding free energies. In SIE, molecular mechanics, the contribution of force field-based interaction energy is complemented by the continuum model of solvation. The implicit solvation model used in MMGBSA and MMPBSA is comprised of parameters that are fitted to the experimental solvation free energies of small molecules, and therefore may be neither sufficient nor optimal for protein-ligand binding free energy calculations. The SIE method is explicitly calibrated on binding affinities in solution by using a diverse training set of 99 protein-ligand complexes. This calibration of the dielectric constant, Born radii, surface tension coefficient, and enthalpy-entropy compensation scaling factor can be perceived as a rough treatment of entropy–enthalpy compensation (Naïm *et al.* 2007, Cui *et al.* 2008). Some studies have shown that SIE predicts binding free energies with high accuracy (Naïm *et al.* 2007, Cui *et al.* 2008, Wang *et al.* 2009, Sulea *et al.* 2011, Sulea *et al.* 2012).

2.3 Ligand properties

Receptors and enzymes can differentiate their ligands and substrates through the recognition of shape and physicochemical properties. This recognition process is vital in virtually all biological functions. However, the recognition of chemical features, shapes, and changes in shape is essential not only in the function and understanding of biochemical reactions but also in the methods used in computational biosciences. As already mentioned, many drug discovery methods and evaluations of compounds rely on the idea that similar compounds have similar properties, and thus, presumably, similar biological functionality (Johnson *et al.* 1989, Patterson *et al.* 1996). These similar molecular properties include the structural and physicochemical properties of molecules (Kortagere *et al*. 2009, Nicholls *et al*. 2010).

2.3.1 Shape

Shape is a fundamentally important molecular feature that often determines the fate of a compound in terms of its molecular interactions with a target protein (Kortagere *et al.* 2009). Shape complementarity between the ligand and the target protein is also an important feature when novel drug molecules are discovered or designed. Molecular shape plays an important role in many approaches related to computational biochemistry, such as the discovery and optimization of lead molecules, the prediction of poses in molecular docking, the search for structure-activity relationships, the comparison of molecular similarities, the design of small molecule libraries, the clustering of similar molecules, and the refining of crystallographic structures (Kortagere *et al.* 2009,

Nicholls *et al.* 2010, Maggiora *et al.* 2014). In drug discovery, both global shape matching and partial substructure shape matching have proved useful; however, molecules are more than just shape they are also volumes and surfaces (Nicholls *et al.* 2010, Maggiora *et al.* 2014).

In shape-based VS, the problem is that validation results are sensitive to the composition of the test data set. Because many drugs are topologically similar, the test set may not be variable enough in terms of molecular scaffolds. If this is the case, then the topological methods will find only similar hits and return very limited scaffolds (Huang *et al.* 2006, McGaughey *et al.* 2007, Jain and Nicholls 2008, Kirchmair *et al.* 2009). Because shape is essential in VS, the sampling of the conformational space to a sufficiently fine degree is crucial (Boström 2001, Sastry *et al.* 2013). In addition, in some VS methods, the description of atoms is not specific enough to capture the subtleties of the properties that are responsible for the activity, such as distinguishing between aromatic and aliphatic hydrophobic areas (Sheridan and Kearsley 2002).

2.3.2 Electrostatics

In addition to shape, electrostatics are particularly important because of their long range influence on polar or charged molecules, including water, aqueous ions, and amino acids, 11 of which are charged or polar in neutral solution (Dong *et al.* 2008). Therefore electrostatics are important in determining the structure, motion, association rate, function, and binding affinity of a wide range of biological molecules (Dong *et al.* 2008).

In ligand binding, energetically favorable charge–charge interactions are formed between the compound and the target, as well as specific salt bridges and hydrogen bonds; however, simultaneously energetically unfavorable desolvation, i.e. the displacement of the bound waters on the surfaces of the ligand and the target protein must occur to allow close binding (Sheinerman *et al.* 2000, Sheinerman and Honig 2002, Dong *et al.* 2008). Regarding the electrostatics, the balance of these two aforementioned energetic contributions determines the binding affinity. Computational simulations can provide atom-level information about the energetics and dynamics of protein-ligand complex structure and interactions; however, many models of biomolecular solvation and electrostatics compromise between the accuracy and the computational expense (Dong *et al.* 2008, Genheden and Ryde 2015).

2.3.3 Properties of drugs

The drug-likeness of a compound indicates the similarity of its features to the features of known drugs. Lipinski's rule of five defines guidelines that are commonly used to evaluate the drug-likeness and oral bioavailability of a compound. According to Lipinski's rule of five, the drug-like molecule may violate at most one of the following rules: molecular weight is lower than 500 g/mol; octanol/water partition coefficient (logP) is lower than five; the number of hydrogen-bond donors is no more than five; and number of

hydrogen-bond acceptors is no more than 10 (Lipinski *et al.* 1997). Lipophilicity and molecular weight are often increased when the affinity and selectivity of the drug candidate are improved, whereupon it may be challenging to maintain the drug-likeness. Hence, in lead molecule discovery, a rule of three, which favors smaller compounds and leaves more room for compound modification, has been proposed. The rule of three suggests that molecules having a molecular weight lower than 300 g/mol, a logP lower than three, a number of hydrogen-bond donors, acceptors, and rotatable bonds no more than three, and a polar surface area less than 60 are lead-like (Congreve *et al.* 2003).

Absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties of a molecule describe the characteristics and behavior of a pharmaceutical compound within an organism. Predictive tools used for the accurate assessment of pharmacokinetic, pharmacodynamic, metabolic, and toxicological properties in the early stages of the drug discovery process are highly useful because violations in these areas are important source of costly late stage failures in drug development (van de Waterbeemd and Gifford 2003, Tao *et al.* 2015). As part of the effort to develop tools to assess ADMET properties, novel computational methods are introduced and old methods are continuously improved. The physicochemical properties of a molecule have an important effect on the pharmacokinetic, pharmacodynamic, and metabolic fate of a compound in the body. Therefore predicting these properties is essential in computational ADMET methods. Most ADMET models are based on experimental knowledge or statistical analyses of the relationships between molecular descriptors and the feature. In general, ADMET prediction methods are either ADMET filters, which are used to trim molecular databases, or ADMET models, especially pharmacophore or QSAR models, which are used to evaluate compounds. The molecular modeling of metabolizing cytochrome P450 enzymes is also a popular method in ADMET prediction (Güner and Bowen 2013, Tao *et al.* 2015).

Pan assay interference compounds (PAINS) appear as frequent hitters (or promiscuous compounds) in many different biochemical assays (Baell and Holloway 2010). PAINS interfere assays in multiple ways. For example, some PAINS compounds are fluorescent or strongly colored and may give a positive signal in an assay even without protein. Other compounds may contain residues of metals used in the synthesis of molecules or as reagents in assays giving rise to false signals. Other PAINS may alter protein function or interfere with proteins chemically without localizing specifically with the binding site Often PAINS also interact with many other proteins than the intended target (Baell and Holloway 2010, Baell and Walters 2014, Hu and Bajorath 2014, Dahlin *et al.* 2015). PAINS behavior is related to substructural features that are not identified by filters commonly used to exclude reactive compounds (Hu and Bajorath 2014, Dahlin *et al.* 2015). In commercially available chemical collections that are typically used as academic screening libraries 4.6–11.6% of compounds are problematic (Baell and Holloway 2010). For example, when all compounds from ChEMBL database with direct interactions against human

single protein targets at the highest confidence level with potency measurement types Ki or IC_{50} (activity unit nM accepted and approximate or verbal potency values discarded) were studied, 3.6% of the compounds contain PAINS substructures (Hu and Bajorath 2014). For comparison, for all compounds in ChEMBL with interactions against any target the corresponding percentage is 9.8 (Hu and Bajorath 2014). Although PAINS usually are a poor starting point for medicinal chemistry, some of these compound may interact with a protein in a specific drug-like way, and thus could be optimized through medicinal chemistry (Baell and Holloway 2010, Baell and Walters 2014).

3 AIMS OF THE STUDY

A wide variety of different computational molecular modeling methods exists. Methods differ in their application area, working principles, and usability, having both strengths and weaknesses. In this study, several computational methods are compared for their ability to identify active molecules and predict binding affinities in order to assess their usability for VS and protein-ligand interaction studies. The study also contributes to the understanding of the relative advantages and limitations of these methods.

4 METHODS

The methods used in the original publications that comprise this thesis are summarized in Table 6. The most important methods are briefly discussed in this chapter. Detailed explanations can be found in the original publications.

TABLE 6 Summary of the methods.

Method	Publication
MOLECULAR DATABASES	
DUD	I, II, III
DUD-E	III
ChEMBL	II, IV, V
SPECS	II
Easy-to-synthesize	II
PDB	I, II, III, IV, V, VI
NEGATIVE IMAGE METHODS	
NIB VOIDOO/FLOOD	I, II
Panther	III
MOLECULAR DOCKING	
GLIDE docking	I, II, V
PLANTS docking	II, V
GOLD docking	IV, VI
MOLECULAR DYNAMICS AND BINDING FREE ENERGIES	
MD SANDER (in AMBER 10)	I, IV
MD NAMD	II, VI
MMGBSA/MMPBSA (in AMBER 10)	I, IV
SIE	IV
Prime MMGBSA	V
OTHER METHODS	
3D similarity searching	I, II, III, IV
PHASE pharmacophore and 3D-QSAR	II, V
VS performance evaluation (ROC-AUC, enrichment factor)	I, II, III, IV
EXPERIMENTAL METHODS	
Fluorescence polarization	II
Organic synthesis (performed by collaborators)	II
SMase assay (performed by collaborators)	VI

4.1 Molecular databases

Active and decoy molecules in the directory of useful decoys (DUD) (Huang *et al.* 2006) (I, II, III) and the Database of Useful Decoys: Enhanced (DUD-E) (Mysinger *et al.* 2012) (III) were used to test and validate the VS methods. DUD is extensively used to benchmark VS protocols, especially molecular docking programs. DUD contains 40 protein targets. The decoys in DUD are selected from the ZINC database (Irwin and Shoichet 2005) based on their physicochemical similarity and topological dissimilarity to active ligands. The decoys are assumed inactive, but they may have not been experimentally tested.

DUD-E is an upgraded and rebuilt version of the DUD database. In DUD-E, the number of protein targets has been increased to 102. As in DUD, in DUD-E, the decoy molecules are selected from ZINC, and they have similar physicochemical properties but dissimilar topology compared to the active ligands. Distinct from DUD, in DUD-E, there are 50 decoys for each active ligand, whereas in DUD the corresponding number is 36 decoys. Hence, DUD-E should offer more diverse, less biased, and more challenging molecule sets for benchmarking (Mysinger *et al.* 2012).

Molecules with known, experimentally measured activities were retrieved from the ChEMBL database (Bellis *et al.* 2011) (II, IV, V). The compounds included in ChEMBL are primarily extracted from the medicinal chemistry literature, and they therefore usually are drug-like or lead-like small molecules with full experimental details. To ensure the homogeneity of the experimental data, active molecules are chosen, if possible, from a single research article, or in some cases from a limited number of research articles when a wider variety of the molecules is required.

The SPECS (Specs, The Netherlands, www.specs.net) (II) database is a commercial, small molecule database of screening compounds. SPECS library compounds exhibit the structural characteristics of a biologically active compound, and they meet the criteria of drug-like compounds. Prior to screening, the SPECS database was filtered according to Lipinski's rule of five, which was complemented by allowing the maximum of six rotatable bonds.

An in-house database of easy-to-synthesize coumarin derivatives (II) contains 75 compounds which have been synthesized for experimental testing. The idea is that the compound collection consists of molecules that are synthesizable from cheap starting materials with one step synthesis, excluding possible protecting groups, performed with microwave-assisted organic synthesis in few minutes.

Protein crystal structures were acquired from the Protein Data Bank (PDB) (Berman *et al.* 2000) (I, II, III, IV, V, VI). PDB is a crystallographic database that contains the 3D structural data of large biological molecules, such as proteins and nucleic acids. When crystal structures were selected for the studies presented here, special attention was paid to the resolution of the structure and the bound ligand molecules that affect protein and ligand-binding area conformation.

4.2 Negative image methods

Two slightly different negative image methods were utilized: NIB VOIDOO/FLOOD models and Panther models.

4.2.1 NIB VOIDOO/FLOOD (I, II)

In our earlier studies, we used NIB VOIDOO/FLOOD models of the ligand-binding site of the protein target. The creation of NIB models relied on VOIDOO/FLOOD (Kleywegt and Jones 1994). The VOIDOO program was originally designed for grid-based cavity searching and cavity volume analysis, mainly for crystallographic purposes. The FLOOD program is used to fill the found cavities with solvent molecules or selected atoms.

In the model creation, the solvent accessible surface area probe in VOIDOO was modified so that the allowed area within the ligand-binding site is closer to polar and charged amino acids but slightly further away from the hydrophobic amino acids. This goal was achieved by decreasing the van der Waals radius of electronegative atoms oxygen and nitrogen from 1.6 Å to 1.2 Å and from 1.75 Å to 1.2 Å, respectively, and increasing the radius of carbon from 1.85 Å to 2.25 Å. Atom-centered MMFF94 charges (Halgren 1996) were added to the amino acids in the ligand-binding area. The charges within 2.7 Å distance from each model point were averaged, and the opposite charge was added to the corresponding model point. Protein flexibility was taken into account by utilizing several protein crystal structures or by taking snapshots from MD simulations.

4.2.2 Panther (III)

Based on similar idea as NIB VOIDOO/FLOOD models, a novel Panther algorithm was developed. Panther can be used for cavity searching, cavity volume analysis, and filling the cavities with chosen molecules as VOIDOO/FLOOD. However, this algorithm was designed mainly for accurate ligand-binding area recognition, determination, and filling the cavities to produce models that are better suited to VS. In Panther, the parameters can be adjusted to influence the cavity search process systematically and the model can be fine-tuned to accommodate the features of the ligand-binding site accurately. Panther is discussed more thoroughly in e.g. section 5.2.

4.3 Molecular docking

4.3.1 GLIDE (I, II, V)

The Protein Preparation Wizard in Maestro (Schrödinger LLC, New York, NY) was used to prepare the protein crystal structures for docking. The receptor grid used for docking was generated by GLIDE. GLIDE standard precision (I, II, V) and extra precision (II) modes were used. The scoring function of GLIDE, the *GScore*, is the following:

$$GScore = 0.065*vdW + 0.130*Coul + Lipo + Hbond + Metal + BuryP + RotB + Site$$
(I, II: GLIDE 5.5 and 5.7, respectively)

$$GScore = 0.05*vdW + 0.15*Coul + Lipo + Hbond + Metal + Rewards + RotB + Site$$
(V: GLIDE 5.9)

where *vdW* is van der Waals energy, *Coul* is coulomb energy, and *Lipo* is a lipophilic term, which is a pairwise term in standard precision docking but is derived from the hydrophobic grid potential for extra precision docking. *HBond* is hydrogen-bonding term, *Metal* is metal-binding term, *RotB* is the penalty for freezing rotatable bonds, and *Site* rewards polar interactions in the active site. *BuryP* (in the former equation) is the penalty for buried polar groups whereas *Rewards* (in the latter equation) rewards and penalizes various features, such as buried polar groups, hydrophobic enclosure, correlated hydrogen bonds, amide twists, and so on, covering all terms except those explicitly mentioned.

4.3.2 PLANTS (II, V)

For PLANTS docking (version 1.2), protein structures prepared with the Protein Preparation Wizard in Maestro (Schrödinger LLC, New York, NY) for GLIDE docking were used. In PLANTS docking $PLANTS_{CHEMPLP}$ scoring was used with default parameters. $PLANTS_{CHEMPLP}$ is an empirical fitness function optimized for pose prediction. Its simplified form is the following:

$$PLANTS_{CHEMPLP} = fitnessPLP - (fchem\text{-}hb + fchem\text{-}cho + fchem\text{-}met)$$

where the piecewise linear potential (*fitnessPLP*) is used to model the steric complementarity between the protein and the ligand. For the distance- and angle-dependent hydrogen and metal bonding, terms from ChemScore scoring are considered (*fchem-hb, fchem-cho, fchem-met*). For detailed information about $PLANTS_{CHEMPLP}$ scoring, see (Korb *et al*. 2009).

4.3.3 GOLD (IV, VI)

For GOLD docking, hydrogen atoms were added to the protein structures using TLEAP in ANTECHAMBER 1.27 (Wang *et al.* 2005). The *GoldScore* fitness function is the original and default scoring function provided with GOLD in versions 5.0 and earlier (GOLD 4.1 in IV; GOLD 5.0 in VI). The *GoldScore* fitness function is the following:

$$GoldScore = S(hb\_ext) + 1.375*S(vdW\_ext) + S(hb\_int) + S(vdW\_int)$$

where *S(hb_ext)* is the protein-ligand hydrogen bonding score (external), *S(vdW_ext)* is the protein-ligand van der Waals score (external), *S(hb_int)* is the score of intramolecular hydrogen bonding in the ligand (internal), and *S(vdW_int)* is the score of intramolecular straining in the ligand (internal). The multiplication of the external van der Waals score by a factor of 1.375 yields an empirical correction that encourages protein-ligand hydrophobic interactions. Internal hydrogen bonding in the ligand is an optional term which was not taken into account in the scoring.

4.4 Molecular dynamics simulations

4.4.1 SANDER (I, IV)

In articles I and IV, MD simulations were used to create protein-ligand complex conformers for binding free energy calculations. For the MD simulations, the starting conformers were generated using the similarity searching method, SHAEP (I, IV), GLIDE docking (I), and GOLD docking (IV). The charges for the ligands were derived from AM1-BCC (Jakalian *et al.* 2000) available in ANTECHAMBER. TLEAP in ANTECHAMBER (Wang *et al.* 2005) was used to create the force field parameters for the protein (ff03, Duan *et al*. 2003) and the ligand (gaff, Wang *et al*. 2004), add hydrogens, and solvate the protein-ligand complex with a rectangular box of transferable intermolecular potential three-point water molecules (TIP3P3, Åqvist 1990) 14 Å in all directions.

The MD simulations were run with the SANDER algorithm distributed in the AMBER 10 package (Case *et al.* 2008). The system was first minimized using the conjugate-gradient method for 1000 steps without restraints. This was followed by an equilibration step at constant volume by allowing the system to heat from 100 K to 300 K for 1000 steps with NMR restraints. The production simulation with a 2 fs time step was run without restraints for 20,000 steps (simulation time 40 ps in I) or 256,000 steps (simulation time 512 ps in IV) at a constant pressure controlled by isotropic position scaling. The temperature was maintained with the Berendsen thermostat (Berendsen *et al.* 1984). Electrostatics were treated with the Particle-Mesh Ewald (PME) method (Darden *et al.* 1993, Petersen 1995), and a cutoff value of 12 Å was employed for non-bonded

interactions. The equilibration step and the production simulation were run under periodic boundary conditions. The SHAKE algorithm (Ryckaert *et al.* 1977) was used to restrain bonds involving hydrogen atoms.

4.4.2 NAMD (II, VI)

In article II, a ligand-free protein structure was simulated to include protein flexibility for NIB VOIDOO/FLOOD modeling. In article VI, MD simulations were used to study the binding properties of ligand analogs to their target protein. TLEAP in ANTECHAMBER (Wang *et al.* 2005) was used to create the force field parameters for the protein (II, VI) (ff03, Duan *et al*. 2003) and for the ligand (VI) (gaff, Wang *et al*. 2004), to add hydrogens, and solvate the protein-ligand complex with a rectangular box of transferable intermolecular potential three-point water molecules (TIP3P3, Åqvist 1990) 13 Å in all directions.

Energy minimizations and MD simulations were performed using NAMD 2.6 (Phillips *et al.* 2005). First, the water molecules and amino acid side chains were minimized with a conjugate gradient algorithm for 15,000 steps, while the alpha carbons were kept in place with the harmonic force of 5 kcal/(mol*Å^2). Second, 15,000 step minimization was performed without constraints. In the third phase, the simulation was run for 360 ps with a 2.0 fs time step at a constant temperature (300 K) and pressure (1 atm). The water molecules and amino acid side chains were allowed to move, while the alpha carbons were restrained as in the first energy minimization step. Finally, in the fourth step, the simulation was performed for 2.4 ns with a 2.0 fs time step (II) or for 7.2 ns with a 1.0 fs time step (VI) at a constant temperature (300 K) and a constant pressure (1 atm) with the Langevin-Piston method. In the MD simulations, a cutoff value of 12 Å was used for the van der Waals interactions, and the electrostatics were treated with the Particle Mesh Ewald (PME) method (Darden *et al.* 1993, Petersen 1995). The simulations were performed under periodic boundary conditions.

4.5 Binding free energy calculations

4.5.1 MMGBSA and MMPBSA (I, IV)

In MMGBSA and MMPBSA, the binding free energies (ΔG_{bind}) can be estimated according to the free energies of three averaged terms for the ligand, the receptor, and the complex over a set of MD snapshots by using the following equation:

$$\Delta G_{bind} = G_{comp} - G_{rec} - G_{lig}$$

where comp is complex, rec is receptor, and lig is ligand. The free energy term for each of the reactants is estimated by using the following equation:

$$G = E_{MM} + G_{solv} - TS_{solute}$$

where E_{MM} is the molecular mechanics contribution in vacuum consisting of the sum of electrostatics, the internal, and the van der Waals energies; G_{solv} is the contribution of solvation free energies as the sum of polar and nonpolar solvation free energies; T is the temperature; and S_{solute} is the solute entropy.

The polar solvation energies were calculated with the GB and PB approaches implemented in AMBER 10 (Case *et al.* 2008). In article II, the GB model parameters developed by Tsui and Case (IGB1, Tsui and Case 2000) were used. In article IV, three different GB models available in AMBER 10 were tested: IGB1, and two GB models developed by Onufriev and co-workers (IGB2 and IGB5, Onufriev *et al*. 2004).

The atomic cavity radii and charges were taken from the corresponding MD topology files. The snapshots were extracted at either 0.4 ps intervals (yielding 100 structures in I) or 4 ps intervals (yielding 128 structures in IV). The dielectric constants of 1 and 80 were used for the interior and the exterior of the molecules, respectively. The hydrophobic contribution to the solvation free energy was estimated by calculating the solvent accessible surface area using Molsurf (Connolly 1983) with a probe radius of 1.4 Å. The surface tension constant γ was set to 0.0072 kcal/mol/Å^2. The estimation of entropy is usually done with a normal mode analysis of the vibration frequencies, but because of the high computational cost of the analysis, the effect of entropy was neglected in the calculations.

4.5.2 SIE (IV)

The SIE method (Naïm *et al.* 2007) is analogous to MMGBSA and MMPBSA. In the SIE calculations (IV), 128 extracted MD snapshots (with 4 ps intervals) were used. SIE calculations can be performed using the software SIETRAJ, which employs the following equation:

$$\Delta G_{bind}^{calc}(\rho, D_{in}, \alpha, \gamma, C) = \alpha[E_{inter}^{C}(D_{in}) + \Delta G_{bind}^{R}(\rho, D_{in}) + E_{inter}^{vdw} + \gamma \times \Delta SA\ (\rho)] + C$$

where E_{inter}^{C} and E_{inter}^{vdw} are the intermolecular Coulomb and van der Waals interaction energies in the bound state, respectively. ΔG_{bind}^{R} is the electrostatic contribution of the solvation free energy to binding. The ΔSA term is the change in molecular surface area upon binding. The coefficients ρ (AMBER van der Waals radii linear scaling coefficient), D_{in} (solute internal dielectric constant), α (fitting coefficient), γ (molecular surface area linear scaling coefficient), and fitting constant C were optimized with the set of 99 protein-ligand complexes (Naïm *et al.* 2007). The optimized values are ρ = 1.1, D_{in} = 2.25, α = 0.1048, γ = 0.0129 kcal/mol/Å2, and C = -2.89 kcal/mol.

4.5.3 Prime MMGBSA (V)

The Prime MMGBSA can be used to calculate the ligand-binding energies for protein-ligand complexes. Prime is part of the Schrödinger Maestro package (Schrödinger, LLC, New York, NY). For Prime MMGBSA, the starting conformers were generated by GLIDE docking. The default settings in Prime version 3.0 were used.

Prime MMGBSA with the VSGB-2.0 solvent model has several empirical corrections (Li *et al.* 2011). For example VSGB-2.0 includes further terms to model physics-based empirical corrections for modeling the directionality of hydrogen bonding interactions, and pi stacking interactions. A special term is used to account for the internal hydrogen bonds in protein structures.

4.6 3D similarity searching (I, II, III, IV)

SHAEP (Vainio *et al.* 2009) was used in the similarity searching. SHAEP can perform similarity comparisons based on either only the shape of the molecules (*onlyshape* option) or by comparing both the shape and the electrostatics of the molecules (the default option). In addition, the weighting of shape and electrostatics can be set (espweight option).

4.7 Pharmacophore modeling and 3D-QSAR (II, V)

Pharmacophore models with the 3D-QSAR option were built in the PHASE module version 3.3 (II) or version 3.5 (V) of the Schrödinger software (Schrödinger, LLC, New York, NY). The PHASE program is suitable for structure alignment, pharmacophore modeling, activity prediction, and 3D database searching. PHASE can use pre-created and pre-aligned conformers, or it can perform conformational sampling prior to model generation and align newly created conformers (II, V). PHASE identifies common pharmacophore hypotheses that represent the mutual characteristics of 3D ligands, which are important for binding, by using a tree-based partitioning technique that groups similar pharmacophores. Pharmacophore hypotheses can then be combined with known activity data to create a 3D-QSAR model (II, V) that recognizes the features of molecules that govern activity (Dixon *et al*. 2006a, Dixon *et al*. 2006b).

4.8 Metrics for method performance evaluation (I, II, III, IV)

The performance efficiency of the VS method can be graphically illustrated by plotting the true positive rate of active ligands (y-axis) against the false positive rate of decoy molecules (x-axis) in a receiver operating characteristics (ROC) curve (Metz 1978, Hanley and McNeil 1982). Traditional ROC curves are drawn on a linear scale. However, when the beginning of the curve, that is, the early enrichment of active molecules, is highlighted, the curve can be drawn on a logarithmic scale. The numerical summary of the ROC curve is given as an area under the curve (AUC) value. AUC values denote the probability of a randomly selected molecule to be an active compound. Completely random sampling yields an AUC value of 0.5, whereas a perfect VS method would produce an AUC value of 1.0.

The early enrichment factors are calculated in two different ways: the ratio of ligands to molecules in a subset (n) divided by ratio of all ligands and all molecules (*EFn%*), or as a true positive rate when 1% of the decoy molecules have been found ($EF1\%_{DEC}$). The following formulas are used:

$$EFn\% = (Ligs_{n\%}/Mols_{n\%})/(Ligs_{all}/Mols_{all})$$

$$EF1\%_{DEC} = (Ligs_{1\%DEC}/Ligs_{all})*100$$

where $Ligs_{n\%}$, $Mols_{n\%}$, $Ligs_{all}$ and $Mols_{all}$ are the number of the ligands in the top n% of the screened compounds, the number of the molecules in the top n% of the screened compounds, the total number of the screened ligands, and the total number of the screened molecules, respectively. $Ligs_{1\%dec}$ is the number of ligands when 1% of the decoy molecules have been found and $Ligs_{all}$ is the total number of the screened ligands.

4.9 Visualization

Figures presenting protein structures or protein-ligand interactions were generated by using BODIL v. 0.81 (Lehtonen *et al.* 2004), MOLSCRIPT v. 2.1.2 (Kraulis 1991), and RASTER3D (Merritt and Bacon 1997).

5 RESULTS

5.1 Virtual screening of phosphodiesterase type 5 inhibitors (I)

NIB modeling had been previously tested for shape-based screening (Virtanen and Pentikäinen 2010). Here, cyclic nucleotide phosphodiesterase type 5 (PDE5) was used as model target for VS using NIB VOIDOO/FLOOD models that were seasoned with electrostatic properties in addition to shape. Furthermore, the effect of the added electrostatics was also studied in an ensemble of nuclear hormone receptors. The charges were averaged from the amino acids surrounding the ligand-binding area, and then the opposite charges were added to the corresponding model points. The study used two PDE5 crystal structures: one was crystallized with tadalafil and the other was crystallized with sildenafil. The studies were conducted either by utilizing only one of the structures or by combining the results. For the validation PDE5-specific active ligands from the DUD database and two decoy molecule sets were used: PDE5-specific decoy molecules from DUD (referred to here as PDE5 decoys); and other active DUD ligands were used as decoy molecules for PDE5 (referred to here as PDE5/DUD decoys). In addition, the effect of post-processing on the early enrichment of the VS results was explored with MD/MMGBSA.

5.1.1 Shape-based screening

Ligand shape-based screening. Neither PDE5 inhibitor, tadalafil nor sildenafil (I, Fig. 2), could efficiently identify the active molecules in the validation test sets. The AUC values of the sildenafil-based searches with both decoy sets were below 0.50 (I, Table 1), which means that random picking identifies a greater number of active molecules. For the tadalafil-based searches, the AUC values were slightly better (I, Table 1: 0.67 ± 0.04 for the PDE5 decoys and 0.58 ± 0.04 for the PDE5/DUD decoys). Combining the ligand structure data on sildenafil and tadalafil marginally weakened the results of both decoy sets (I, Table 1: 0.64 ± 0.04 for the PDE5 decoys, and 0.57 ± 0.04 for the PDE5/DUD decoys).

NIB shape-based screening. NIB VOIDOO/FLOOD models were created for both PDE5 crystal structures. The smaller NIB solvation model (NIB-SM1) was based on tadalafil-bound crystal structure, whereas the larger NIB model (NIB-SM2) utilized a sildenafil-bound crystal structure (I, Fig. 2). Shape comparison of the NIB VOIDOO/FLOOD models produced higher AUC values than those for the ligand shape-based screening (I, Table 1). The smaller NIB-SM1 produced AUC values that were moderately higher than those for the larger NIB-SM2 (I, Table 1: 0.72 ± 0.04 vs. 0.65 ± 0.04 for PDE5 decoys and 0.75 ± 0.04 vs. 0.71 ± 0.03 for PDE5/DUD decoys). Unlike the ligand shape-based screening, the combinatory usage of the NIB VOIDOO/FLOOD models did not weaken the AUC values (I, Table 1), but it identified a greater number active molecules in the later phases (I, Tables 2 and 3).

5.1.2 Effect of electrostatics

Docking. The advantage of docking is that it takes into consideration both the shape and the electrostatics within the target protein ligand-binding area. The GLIDE docking produced weaker AUCs when the tadalafil-PDE5 crystal structure was utilized instead of sildenafil-bound protein structure (I, Table 2: 0.68 ± 0.04 vs. 0.78 ± 0.04 for the PDE5 decoys and 0.59 ± 0.04 vs. 0.78 ± 0.04 for the PDE5/DUD decoys). The usage of both crystal structures in docking produced the highest AUC value 0.80 ± 0.03 (I, Table 2: PDE5 decoys). The docking showed good results for both decoy sets when the number of active hits (I, Table 3) and the level of enrichment (I, Table 4) were inspected.

Ligand-based screening. The addition of electrostatic information notably improved the ligand-based VS in the tadalafil-based or the combined search using the PDE5/DUD decoys (I, Table 1 vs. Table 2; Tables 3 and 4). However, the electrostatics enhanced only marginally the results of the PDE5-specific decoys (I, Table 1 vs. Table 2). The sildenafil-based searches with electrostatics remained ineffective in both decoy sets (I, Table 2). Using both PDE5 inhibitors in combination did not improve the overall results of the VS (I, Table 2).

NIB screening. The addition of electrostatics improved the AUC values of the NIB screening with the tadalafil-like NIB-SM1 and the combined NIB model search with the PDE5/DUD decoys (I, Table 1 vs. Table 2). However, the addition of electrostatics marginally decreased the results of both solvation models with the PDE5-specific decoys and the NIB-SM2 results with the PDE5/DUD decoys (I, Table 1 vs. Table 2). The usage of both PDE5 inhibitors in combination did not improve the overall results of the VS (I, Table 2). Both the number of active hits and the enrichment in later percentages benefitted from the addition of electrostatics, whereas the early results were actually weakened by the addition of electrostatics (I, Tables 3 and 4).

Overall, according to the AUC values, NIB VOIDOO/FLOOD and the ligand-based screening benefitted from the added electrostatic information, especially with the PDE5/DUD decoy set (I, Table 2). Different weightings of the shape and the electrostatics were tested, but the best results were achieved

when they contributed equally (data not shown). Docking achieved the highest enrichments and identified the most active molecules (I, Tables 3 and 4).

Effect of electrostatics on other targets. The effect of the addition of electrostatic information was also studied in targets with less polar ligand-binding sites than that of the PDE5. These targets have been tested previously for shape-based NIB VOIDOO/FLOOD modeling (Virtanen and Pentikäinen 2010). The docking results showed that it was somewhat less efficient than the ligand-based and NIB screening were (I, Table 6). However, with the estrogen receptor α (ERα) antagonist conformer, the docking produced high early enrichments (I, Table 6). The addition of electrostatics in the ligand-based VS improved the results considerably in all studied cases compared to the ligand shape-based results (I, Table 6). In general, the addition of electrostatic information to the NIB VOIDOO/FLOOD models produced either comparable or slightly better results than the models without charges; however, in some cases, the early enrichment was substantially weakened (I, Table 6).

5.1.3 Post-processing of the virtual screening results

The top 5% of the VS results were post-processed with MD/MMGBSA to improve the early enrichment further. Short (40 ps) MD simulations were run to optimize ligand poses at the ligand-binding site. The re-ranking according to the favorability of the interactions within the protein-ligand complex was evaluated using the MMGBSA calculations.

Rescoring of the docking-based screening. The docking with GLIDE produced good AUC values and very early enrichments. The post-processing with MD/MMGBSA did not improve the AUC values or the early enrichments of docking screening (I, Tables 2 and 4). Actually, after the post-processing the early enrichment was weakened considerably in both decoy sets (I, Table 4).

Rescoring of the ligand-based screening. Although the AUC values did not indicate any change in the performance of the ligand-based screening when its results were post-processed with MD/MMGBSA (I, Table 2), the relative number of the active hits (I, Table 3) and the early enrichment (I, Table 4) improved nicely. The improvement can be visually inspected in the top 5% of the ROC curves (I, Fig. 3). In particular, the effect of post-processing is seen in a 20-fold enrichment in the top 0.5% of the ranked results of the PDE5/DUD decoy set (I, Table 4). Improvement was not as notable in the PDE5 decoy set.

Rescoring of the NIB screening. Similar to ligand-based screening, the AUC values in the results for the NIB VOIDOO/FLOOD did not show improvement after post-processing with MD/MMGBSA (I, Table 2). However, the improvement is visible in the top 5% of the ROC curves (I, Fig. 3), in the number of active hits (I, Table 3), and in the top enrichment (I, Table 4). The highest enrichments were 39-fold in the top 0.5% (i.e., 20% of ligands) and 24-fold in the top 1% of the PDE5/DUD decoys (I, Table 4).

Overall, it can be concluded that ligand-based and NIB VOIDOO/FLOOD screening benefitted from post-processing with MD/MMGBSA. However, the post-processing was not profitable in the docking process.

5.2 Virtual screening and identification of selective estrogen receptor modulators (II)

In this study, the effectiveness and feasibility of ligand-based common pharmacophore hypothesis, followed by 3D-QSAR and the structure-based methods, molecular docking and NIB screening, were compared. ERα, a widely applied drug discovery target, was used as a model structure. Validation of the methods relied on two datasets (DS): DS1 and DS2. DS1 contains 101 chemically diverse selective estrogen receptor modulators (SERMs) with ERα inhibitory activity IC_{50} values ranging from 0.2 nM to 10,000 nM retrieved from the ChEMBL database. DS2 consists of ERα antagonists and corresponding decoy molecules from the DUD database. In the study, a workflow for rapid identification of SERMs was suggested. The workflow was applied to the screening of the commercial Specs molecule database and our own compound collection of easy-to-synthesize molecules. The top-ranked compounds then were tested experimentally.

5.2.1 Efficiency of pharmacophore modeling and 3D-QSAR

The best pharmacophore hypothesis followed by building of 3D-QSAR model (II, Fig. 1: APRRR-223) was selected based on various statistical grounds (R^2 = 0.923, standard deviation SD = 0.317, Fischer significance F = 154.7, chance correlation P = 4.47×10^{-44}). The survival numbers of active and inactive ligands in the best model were 43.144 and 42.517, respectively. When DS1 was studied, the selected model identified 61 of the most active ligands from 101 ligands. However, two active ligands with pIC_{50} > 9 (ChEMBL241301 and ChEMBL391910) were not identified.

The 3D-QSAR model was validated by randomized test set from DS1 molecules. The validation statistics (Q^2 = 0.822, RMSE = 0.431 and R^2 = 0.870) showed the accuracy of the model in predicting the ERα activity in the test set. The 3D-QSAR model also was tested using DS2. Compulsory matching of all five pharmacophore features in the hypothesis showed that from 39 active ligands and 1395 decoys, the hypothesis recognized 16 of the most active antagonists without selecting any decoys. When the utilization of only four pharmacophoric features was required, all 39 active ligands were recognized, but also more than 600 decoys (43%) were identified. These results showed that all five chemical features included in APRRR-223 hypothesis are important.

According to the validation, the pharmacophore and 3D-QSAR model preferentially selected only the highly active ligands from both DS1 and DS2, which indicates that it could be used for the identification of SERMs from a database in the VS study.

5.2.2 Efficiency of molecular docking

Two docking methods were used: GLIDE and PLANTS. Both docking methods employed the original protein crystal structure (PDB: 3ERT). In addition, a slightly relaxed protein structure used in NIB VOIDOO/FLOOD modeling (see below) was also tested in docking, but because the minimization of the protein did not improve the docking results, it is not discussed here.

The validation of the docking methods was done by studying the correlation between the ERα activity of DS1 ligands and the corresponding docking scores (II, Fig. 2). The faster GLIDE SP docking was considerably better than the extensive GLIDE XP docking ($R^2 = 0.638$ vs. $R^2 = 0.230$, respectively). In the case of the PLANTS docking with CHEMPLP scoring, the regression model yielded the correlation coefficient $R^2 = 0.639$. To improve the accuracy of the docking, a hybrid regression model was developed by utilizing the normalized docking scores of both GLIDE SP and PLANTS CHEMPLP, which enhanced the correlation coefficient to $R^2 = 0.787$. The visual inspection of the top-ranked poses of GLIDE SP and PLANTS CHEMPLP showed reasonable binding modes.

In addition, the ability of the docking methods to separate active molecules from inactive molecules was studied in both the DS1 and DS2 molecule sets and then evaluated with ROC curves and AUC values (II, Fig. S2). The AUC values showed better performance in GLIDE SP docking than in PLANTS CHEMPLP docking in both DS1 and DS2 (II, Table 1: 0.91 ± 0.01 vs. 0.65 ± 0.01 for DS1 and 0.91 ± 0.01 vs. 0.73 ± 0.01 for DS2). The hybrid docking model yielded AUC values of 0.79 ± 0.01 for DS1 and 0.86 ± 0.01 for DS2.

Overall, the GLIDE SP docking identified active molecules efficiently in DS1 containing larger number of SERMs and in DS2 with a limited amount of antagonists. Even though the efficiency of separating the active ligands over decoys with PLANTS CHEMPLP was slightly lower than with GLIDE SP, the results of PLANTS showed improvement in the separation of lower activity molecules from the datasets when it was used in combination with GLIDE. Therefore, the hybrid docking model showed a balanced level of screening performance in separating the active molecules from decoys in both DS1 and DS2.

5.2.3 Efficiency of negative image-based screening

Ligand-binding area flexibility was included in the NIB VOIDOO/FLOOD models by using protein conformers derived from MD simulation snapshots, which resulted in 21 models. All NIB VOIDOO/FLOOD models were screened against DS1 and DS2 by using SHAEP, which yielded AUC values ranging from 0.01 to 0.97 in DS1 and from 0.38 to 0.85 in DS2. These results were obtained using the NIB VOIDOO/FLOOD search with the charge distance of 1.8 Å and an equal shape and electrostatic weighting of 50%. In addition, the binding poses of DS1 and DS2 ligands in various models were studied. The results suggested that model 1 (II, Fig. 3: a slightly relaxed crystal structure) was the most promising; thus, it was selected for further evaluation and optimization.

The experimentally most active compounds tend to have good balance in their electrostatic and shape contribution indicating that shape plays a crucial role in ligand binding but the electrostatic interactions lead to high affinity. The balance of the shape and electrostatics for the chosen NIB VOIDOO/FLOOD model 1 in VS was validated by varying the degree of electrostatic weighting from 0% to 100% at intervals of 10% (II, Table 1). These results clearly indicated that the shape alone or particularly low or high electrostatic weighting was not beneficial in effectively distinguishing active molecules from decoys. The most optimal electrostatic weighting was around 50% (II, Table 1 and Fig. S3).

5.2.4 Identification of potential selective estrogen receptor modulators

Using combination of four above presented methods screening identified no potential high affinity molecules ($pIC_{50} > 7.0$) from the pre-filtered commercial Specs molecule database containing about 100,000 drug-like molecules. However, if only both docking methods were considered, five potential molecules (II, Table 2) appeared. When these molecules were experimentally tested, only one (II, Table 2: molecule S4) showed activity ($pIC_{50} \approx 6.6$) within the concentration range of 0.0007 nM to 10 000 nM. Additionally, the developed workflow was used to screen and analyze the small virtual library of easy-to-synthesize molecules. The experimental testing showed that the top five molecules from the virtual library had pIC_{50} values from 5.5 to 6.5 (II, Table 2).

Based on the structural characteristics, the Specs molecule is likely an antagonist while the easy-to-synthesize molecules are agonists. Notably, the computational predictions of experimental activity were relatively accurate (II, Fig. 4). Also noteworthy is that among the presumably inactive decoy molecules, the workflow methods identified a highly active ligand ($pIC_{50} = 8.0$). Moreover, some interesting DUD decoy molecules were repeatedly identified, but unfortunately they are not commercially available and thus cannot be tested.

5.3 Novel Panther method (III)

Molecular docking is commonly used in protein structure-based VS. This study presents Panther, a novel, simple, fast, and efficient multipurpose docking tool. Similar to molecular docking, Panther utilizes structural information and requires a high-quality protein crystal structure. In Panther, the ligand-binding area of the protein is described as a simple atomistic shape-electrostatic model. Molecular databases can then be compared to the model by using the similarity search algorithm SHAEP (Vainio *et al.* 2009), which is used in ligand-based VS methods. In this study, the performance of Panther was evaluated by screening DUD and DUD-E datasets for nine protein targets that are familiar from our previous studies (Virtanen and Pentikäinen 2010, Niinivehmas *et al.* 2011). The results were also compared to previous results of the NIB VOIDOO/FLOOD screening and the results collected from the literature.

5.3.1 Generation of the Panther models

To limit the search area, in the first phase, the approximate location of the binding cavity is defined by allocating point(s) in the presumed cavity. The protein and possible cofactor atoms then are divided into spherical sectors to select the amino acids lining the binding site (III, Fig. 1A). More than one center point can also be used to accommodate the various shapes of binding cavities. In this case, calculations are repeated for each of the used centers.

For cavities on the protein surface, prominent shapes may be included in the amino acids lining the binding site even though they are not part of the binding pocket. To avoid this, only lining atoms with a defined radius (e.g. 4 Å) from other lining atoms are included. A similar problem may occur in protein structures containing multiple chains or subunits; therefore, only amino acids from the closest protein chain are routinely safe to use. If necessary, the amino acids lining the binding site can be either included or excluded by the user.

When the lining of the ligand-binding site has been obtained, model atoms are placed around the polar/charged atoms at a specified angle, geometry (e.g. planar or tetrahedral), and distance of both the lining amino acids and the cofactors (III, Fig. 1B). Typically, the charges for the model atoms are defined as opposite the charge of the lining polar/charged atom, or they are taken from an additional input file (e.g. AMBER 12 force field charges).

When hydrogen atoms are not included in the protein structure file, only one model that correctly describes the possible interactions of amino acids with hydroxyl groups cannot be defined. Similarly, if information about the protonation of histidines is lacking, all three alternative protonation states need to be considered. When the location of the hydrogen cannot be defined by studying the donors and acceptors in the environment, multiple alternative models are generated. If hydrogen atoms are present in the protein PDB-file, their positions are considered fixed and only one model is generated.

Ligand binding pockets may include water molecules that are either critical for ligand binding or replaced by a ligand. Accordingly, the deletion or usage of such water molecules results in different cavity conformers which can be taken into account in the model creation. If a water molecule has hydrogen atoms in the PDB-file, the position of water molecule can be used as is. In addition, if two donors or acceptors are found within the hydrogen bonding distance within an acceptable angle, the water molecule can be taken into account.

After the charged model points are defined, the cavity is filled with non-charged filler atoms to represent the shape of the cavity. The filling can be grid-based, or a face-centered cubic lattice method can be used. Model points overlapping the protein atoms or the created charged atoms or those located too far from the protein are discarded (III, Fig. 1). Finally, various exclusion options are applied to the model points to finalize the model.

5.3.2 Evaluation of the performance of Panther screening

Shape and electrostatic contribution. In general, two Panther models for each of the nine protein targets were created (except one model was created for the mineralocorticoid receptor and three models were created for the ERα). In most cases, both models worked equally well, and thus the model selection was more or less insignificant. However, the performance of the ERα and glucocorticoid receptor models demonstrated the importance of the shape in ligand recognition by showing that because the molecules in datasets are varied, differently shaped models suit different screening molecules. There was also a slight difference in the performance of the PPARγ (peroxisome proliferator activated receptor gamma) model and the PDE5 model, which most likely is caused by variations in the electrostatic points in the models. In the Panther PR-3kba model (a model based on a protein structure with PDB code 3KBA), one charge point clearly blocked another charge point. When this charge point was manually removed, and the root mean square deviation (RMSD) test was repeated, the results improved considerably. This finding indicates that although model building is quick and relatively accurate, the positioning and utilization of the electrostatic points is not trouble-free. Overall, these results showed the importance of both the shape and electrostatics in ligand binding.

In addition, the weighting between the shape and the charge contribution is an important parameter in molecular overlay and similarity searches. However, the usefulness of the electrostatic component is target-protein specific, so protein targets with polar binding sites tend to benefit from the overweighting of electrostatics, whereas hydrophobic targets do not profit from it. In our previous study with PR, the AUC value increased notably (from 0.50 to 0.77) when the electrostatic weighting was raised from 50% to 70% (Virtanen and Pentikäinen 2010). In Panther, almost the same levels of both the AUC value and the early enrichment were reached in the PR screening with 50% electrostatic weighting as achieved using 70% electrostatic weighting previously in NIB VOIDOO/FLOOD model. Similarly, when the electrostatic weighting in Panther was raised to 70%, the same improvement in AUC values occurred in the DUD screening. In addition, the binding of a small molecule to a protein is dependent on the protonation state of the molecule, which in turn depends on the solvent pH value (Shelley *et al.* 2007, Petukh *et al.* 2013, Sastry *et al.* 2013, Urbaczek *et al.* 2014). Thus, because protein-ligand recognition and binding are pH dependent, the preparation process of both the protein and the ligands is important (Shelley *et al.* 2007, Sastry *et al.* 2013, Urbaczek *et al.* 2014). In this study, low energy ligand conformers were created in pH 7.4 ± 0.0, which does not always cover protonation states sufficiently. For example, the original protonation state for the PDE5 ligands sildenafil and vardenafil was reproduced only when a one to two unit tolerance in pH was allowed.

Comparison to NIB VOIDOO/FLOOD and methods from the literature. The novel Panther algorithm outperformed the previously used NIB VOIDOO/FLOOD approach, by producing either better or equal AUC values.

Overall, six of eleven Panther models with DUD screening led to either excellent (>0.9) or very good AUC values (0.9–0.8). Four models showed good AUC values (0.8–0.7), and one model showed a fair AUC value (0.7–0.6) (III, Table 2). When the early enrichments of NIB VOIDOO/FLOOD and Panther were compared, both performed better than the other in an equal number of cases (III, Table 3: four of eleven cases). However, the difference in enrichments was higher for benefit of Panther. In the novel Panther algorithm, model building is fast and rational and a much more flexible treatment of the protein structure is enabled than in the NIB VOIDOO/FLOOD. The downside is that the time required for the screening process increases slightly because the novel models are more precise and have more representative data points. Nevertheless, using a standard PC, the computational time required per ligand was only about 100 ms.

Compared to Panther, traditional molecular docking methods are computationally demanding. When the DUD dataset was screened, eHITS achieved excellent AUC values in five of the ten cases tested (III, Table 2). FlexX and Panther also achieved excellent AUCs in several cases (III, Table 2). One disadvantage of Panther is the high dependence on the generated conformers that often is a compromise between a reasonable number of conformers and the increased computing time. Especially in very flexible molecule sets, the number of conformers sometimes may be too limited, which also limits the performance of Panther. The PDE5 results showed that the choice of protein structure (i.e. acknowledging flexibility) was highly important. Similarly, the importance of the protein structure selection was shown in cyclooxygenase 2, where small changes in the amino acid side chain positioning led to very different Panther models. In addition, the composition of the validation dataset is significant. For example, the ligands that are similar and easily separable from the decoys in the retinoid X receptor alpha DUD set, which makes it an unrepresentative benchmarking set. Hence, all the compared methods performed well in screening retinoid X receptor alpha revealing nothing of the differences between the methods.

In the DUD-E screening with Panther, one of the nine tested models led to excellent AUC values. Five yielded good AUC values, and three achieved fair results (III, Table 4). When the DUD-E set was screened, the performance of Panther was superior to all the compared methods (three docking methods and MMGBSA) in five cases of nine (III, Table 4). In other cases, the AUC values of Panther were comparable to other methods. The average of the AUC values for DUD-E was 0.73, which was lower than the average result of 0.82 for DUD screening (III, Table 2 vs. Table 4). Because the DUD-E benchmarking set is more extensive and more challenging than DUD, the results achieved by the Panther can be considered good. However, all early enrichments in the DUD-E screening were rather poor (III, Table 5). When Panther was used in the DUD-E screening, the problem was that although the overall results were very good, the beginning of the VS was bad (III, Fig. 2). A suboptimal beginning in the screening is undesirable because when the screening procedure is applied to a

huge database, only the top ranked compounds usually are tested further. This shows that high AUC values do not automatically indicate good results.

Reasonability of the docking poses. The size and the flexibility of the co-crystallized ligands varied notably. When the ten best-ranked conformers for 18 studied crystal ligands (four ligands had less than ten conformers) were taken into account, 69% of the poses had an RMSD < 2 Å and thus were considered correct. Of the poses, 14% had a reasonable RMSD ≥ 2 Å but ≤ 4 Å, and 17% had an RMSD > 4 Å (III, Fig. 3). Eleven of the 18 models repeated the crystal structure ligand pose very well, and the top ten poses were generally error free, including four rigid ligands with less than ten conformers. In fact, in two of the 18 tested targets, the theoretically best RMSD of the generated conformers was > 2 Å; therefore, it was not possible to identify the correct pose in those targets. For all other targets, the theoretically best RMSD was < 1 Å. Although the original crystal structure ligand was a long and flexible acid molecule in three of the protein targets, which generally is difficult for both the conformer generation process and the crystal structure repeat, repeating the poses succeeded surprisingly well. Among the tested target structures were two flexible PDE5 ligands: sildenafil and vardenafil, whose binding orientation is more or less C-shaped. The somewhat symmetrical distribution of electrostatic points in the Panther model may guide conformers to flip by 180°. On the other hand, flexible ERα antagonist structures are T-shaped. In ERα Panther models, because the charged areas are clearly distinguishable, the positioning of the ligand conformers in the models was easier, and no bad poses existed among the top ten poses.

5.4 Comparison of binding free energy calculation methods (IV)

In this study, the performance of MMGBSA, MMPBSA, and SIE in VS were assessed. MMGBSA was calculated with the three GB models available in AMBER: IGB1 (Tsui and Case 2000), IGB2 (Onufriev *et al.* 2004) and IGB5 (Onufriev *et al.* 2004). The VS efficiency of the methods was compared, the ability of the methods to predict the binding affinities was evaluated, and the effect of the length of the MD simulation on the results was determined. Five protein targets—aldose reductase 2 (ALR2), β-lactamase (AmpC), human heat shock protein 90 (HSP90), phosphodiesterase type 5 (PDE5), and progesterone receptor (PR)—were chosen to represent variable ligand-binding areas. The DUD dataset and molecules from CHEMBL with known IC_{50} values for targets were used in the VS studies. The protein-ligand complexes were prepared by using two different methods: similarity superimposition of molecules with SHAEP onto crystal structure ligands; and molecular docking with GOLD.

5.4.1 Virtual screening efficiency

ALR2. Overall, the VS results of both GOLD and SHAEP showed very low VS efficiency for ALR2 when the AUC values were considered. Nevertheless, the early enrichments for GOLD were rather good. The best results in VS were achieved by the GOLD-generated complexes treated with MMGBSA with IGB5 parameters (IV, Table 2: 0.65 ± 0.07). In general, the GOLD-generated complexes produced better results than SHAEP generated structures. IGB1 and IGB2 did not succeed in VS with either GOLD- or SHAEP-generated complexes. MMPBSA and SIE performed equally although they were weaker than the results of the IGB5 in the GOLD-generated structures.

AmpC. In AmpC, the VS efficiency with docking was really poor, but ligand-based VS with SHAEP performed considerably better, producing an AUC value of 0.76 ± 0.06 and the enrichment factors clearly higher compared to docking (EF1% = 17.9, EF5% = 6.6, EF10% = 3.8). All post-processing methods clearly improved the VS efficiency in terms of AUC values for the GOLD-generated structures (IV, Table 3). Nevertheless, in general, the SHAEP-generated complexes produced better results than the GOLD-generated structures. In both GOLD- and SHAEP-generated structures, the MMGBSA with the IGB5 parameters produced the highest AUC values (0.82 ± 0.04 and 0.89 ± 0.03, respectively), as well as the highest enrichment factors.

HSP90. The VS results for HSP90 with GOLD were equal to random picking, but SHAEP performed considerably better in terms of the AUC and the early enrichments. The post-processing of the structures generated by GOLD or SHAEP improved the AUC values only in the case of IGB5 (IV, Table 4). In all other methods, the VS efficiency decreased along the MD simulations. Nevertheless, SIE and IGB1 performed better in the early enrichment (IV, Table 4).

PDE5. In VS with PDE5, the AUC values showed that docking performed better than SHAEP, although the early enrichment was better in SHAEP. IGB1 and IGB2 performed comparably according to the AUC values of both GOLD- and SHAEP-generated structures producing also the highest early enrichments (IV, Table 5). In particular, the early enrichment benefitted from the usage of SHAEP to create the starting geometries. The results of the GOLD-generated complexes with MMPBSA and SIE were roughly comparable to the results of IGB2 and IGB2 in terms of AUC values and early enrichments. In SHAEP-generated complexes, the results showed that the performance of MMPBSA and SIE was slightly weak. The MMGBSA with IGB5 parameters produced very poor results compared to the other tested methods.

PR. The performance of both GOLD and SHAEP in VS with PR was poor. In the complexes generated with GOLD, all post-processing methods improved the results significantly (IV, Table 6). According to AUC values the best method was MMGBSA with IGB5 parameters; however, the early enrichment was rather poor. The AUC values produced by MMPBSA were nearly as good as those produced by the IGB5, but the early enrichment was clearly better. The

structures generated by SHAEP produced the highest AUC value for IGB5; however, in the early enrichments the improvement was visible only in 10%.

Identification of highly active molecules. In addition, the ability of the methods to separate highly active molecules ($pIC_{50} \geq 6$) used in binding affinity predictions (see below) from the DUD molecules was investigated (IV, Table 13). For ALR2 there was only one highly active molecule, which was identified within the top 1% only when IGB5 and MMPBSA with the SHAEP generated structures were used (IV, Table 13). SIE and MMPBSA with the GOLD-generated structures found the molecule in the top 5%. Other methods, except IGB2 with the GOLD-generated complex, found the molecule in the top 10%. The somewhat poor overall performance was unexpected because the protein crystal structure used in the studies was originally bound to the same inhibitor. For both HSP90 and PDE5, there were 11 highly active molecules. The results showed very poor efficiency in identifying the active molecules among the DUD molecules (IV, Table 13). For example, the MMGBSA methods could not identify any of the PDE5 molecules within the top 10%. For AmpC, there were 13 highly active molecules. IGB1 and SIE with both GOLD- and SHAEP-generated structures and MMPBSA with a GOLD-generated structure found one to two molecules within the top 1%, whereas other methods did not find any molecules within the top 1%. These methods found two to five of the active molecules in the top 5%. No significant change was observed when the top 10% was scrutinized. Altogether less than half of the molecules were identified. For PR, eight structures were highly active. IGB5 with both tested complex generation methods and SIE with the GOLD-generated complexes identified all the active molecules within the top 5% (IV, Table 13). Moreover, the number of identified active molecules in the top 1% was higher in these methods than in other methods; however, other methods achieved a comparable level in the top 5%. Overall, the results showed that the identification of highly active molecules among the DUD molecules was unsuccessful, except for PR.

5.4.2 Binding free energy calculations and experimental activity

The applicability of the methods to predict the binding affinities of molecules with known IC_{50} values was evaluated. For each protein target, ligands were chosen from a single research article for data uniformity.

ALR2. The ALR2 inhibitor set covered six structurally diverse molecules with a wide range of activities (IV, Table 1, Fig. S1). In general, the SHAEP-generated complexes yielded better results compared to the GOLD-generated structures. Correlations with IGB1, IGB2, and MMPBSA were very good in both complex generation methods and in SIE with SHAEP-generated complexes (IV, Table 7). However, the best correlation, 0.94, was obtained by the GOLD-generated complex with IGB2 parameters at 4 ps (IV, Table 7). Interestingly, the results of the complexes generated by GOLD worsened significantly during the 512 ps simulation, whereas in the complexes made with SHAEP, the correlations remained quite stable throughout the simulation.

AmpC. The molecular dataset for AmpC comprises 26 cephalosporin-derived structures with a wide range of activities (IV, Table 1, Fig. S1). The results showed that the correlation coefficients were negative in almost every studied case (IV, Table 8). Only the starting geometries generated with GOLD and the free energy of binding, which was calculated at the first 4 ps time point, produced a positive correlation; however, these correlations were very poor. Perhaps the changes in protein conformation during the MD simulation favored the weaker binding compounds.

HSP90. The HSP90 inhibitor set contained 14 molecules with a relatively narrow range of activities (IV, Table 1, Fig. S1). The best correlations were obtained by GOLD-generated complexes with MMPBSA; however, in all methods, all the correlation coefficients were very poor (only 0.21 at best; IV, Table 9). IGB5 performed well in the identification of active ligands in VS, but in predicting the binding affinities of molecules, the same approach produced clearly the poorest results.

PDE5. Tadalafil and eight analogs (IV, Table 1, Fig. S1) formed the dataset for PDE5. However with measured activities for several stereoisomers the total number of different structures in the set is 22. Because the differences among molecules are extremely subtle, it is not surprising that the correlations produced by the different methods were only moderate (IV, Table 10). The most efficient in predicting the binding affinities was IGB5 with SHAEP-generated complexes, which produced a correlation of 0.35. When three molecules with four different stereoisomers were considered separately, two showed relatively good correlations (IV, Fig. 4, Table 11: molecules 4a-d, 8a-d). For the third compound, the correlations were very low or negative, which was surprising because the only difference was that an ethyl group in (8a-d) was replaced by a methyl group (in 7a-d) (IV, Fig. S1).

PR. The molecular set for PR contained 12 molecules with rather similar structures and activities (IV, Table 1, Fig. S1). The highest correlations were produced by SIE with GOLD-generated complexes (IV, Table 12). In addition, the IGB2 parameters with the GOLD-generated complexes produced reasonable correlations. In general, the results showed that MMGBSA with IGB1 and IGB2 outperformed both IGB5 and MMPBSA.

5.4.3 Effect of the molecular dynamics simulation length

The effect of the length of the MD simulation on the VS was assessed by calculating AUC values and enrichment factors for structures extracted in different time point (4 ps, 32 ps, 64 ps, 128 ps, 256 ps, 384 ps, and 512 ps). The AUC values showed that the VS efficiency of the methods stayed at a stable level for ALR2, PDE5, and PR throughout the MD simulation (IV, Fig. 3). The longer MD simulations were not beneficial for AmpC and HSP90. In fact, a single snapshot from the beginning of the simulation often showed comparable or even better results than the longer simulations. These results imply that also short MD simulations can be used in VS.

The effect of the MD simulation length on the correlation coefficients of the predicted binding free energies and experimentally measured activities were also studied at the same time points. Contrary to the VS efficiency, the effect of the simulation length varied more in the binding free energy calculations. For example, the ALR2 correlation coefficients in the GOLD-generated complexes decreased along the length of the MD, whereas in the SHAEP-generated complexes the efficiency remained rather stable, suggesting that for some complex generation methods also short simulations are sufficient to predict the binding affinities. For PR, the correlations improved in the GOLD-generated structures according to the length of the simulation, whilein the SHAEP-generated complexes, the results weakened after the first time step, after which they either remained rather stable or improved slowly. At the first 4 ps time point for PDE5 with GOLD-generated structures, the correlations were often fair and then dropped significantly before rising according to the length of the simulation. In structures generated by SHAEP, however, the correlation at the first time point tended to be poor, after which the correlation improved constantly until the end of the simulation. This result in turn suggests that longer simulations may be profitable. For AmpC, only the GOLD-generated starting geometries and the binding free energy calculated at the 4 ps time point produced a positive correlation. At all other time points, the correlation coefficients were negative. Similarly, for HSP90, only few correlations of the GOLD-generated starting geometries with MMPBSA were acceptable, and all were obtained at the simulation length of 128 ps.

5.5 Exploring the virtual screening methods (V)

This study assessed common computational methods, molecular docking, MMGBSA, and pharmacophore modeling complemented with 3D-QSAR, to estimate their ability to predict the activities for cyclic nucleotide phosphodiesterase type 4B inhibitors and to determine whether they could be used to separate highly potent inhibitors from less active molecules. This study used 152 molecules with a pIC_{50}-range of 3.4-10.5, which were obtained from six original studies (named P1-P6). Two protein structures were utilized, one of which had an additional alpha helix, which restricted the binding cavity by acting as a lid.

5.5.1 Virtual screening methods in activity predictions

Molecular docking. Two different docking algorithms, PLANTS and GLIDE, were utilized. In PLANTS docking both used protein structures produced reasonable results although the protein structure with the additional alpha helix (the lid) performed slightly better. When the docking poses were inspected, the ligand poses showed greater variation when they were docked into the protein structure without the lid. The correlations between the docking scores and

measured activities within an article were sometimes very low (e.g. P2 and P5), but were immediately improved when more data were added (V, Table 2). The highest correlation, based on simple numerical comparison, was achieved when articles P1-P3 were used in combination (V, Table 2). Delightfully, the correlation coefficients of the entire ligand set (P1-P6) were reasonable in both protein structures (V, Table 2, Fig. 2). The cross-correlation test, where each of the used articles (P1-P6) was left out one at a time, yielded stable correlations with standard deviations usually below one pIC_{50} unit (V, Table 3). The performance of GLIDE was studied as that of PLANTS, but the results were not successful (V, Table 4).

Prime MMGBSA. The reasonability of the best-docked ligand poses was further estimated by rescoring the results of both docking methods using Prime MMGBSA. Furthermore, the top five output conformers for each ligand in the PLANTS docking were also rescored. In general, the rescoring of PLANTS docking results using Prime MMGBSA worsened the correlations with the experimental pIC_{50} values compared to results achieved using only PLANTS docking without rescoring. The only exceptions were P5 with both protein structures and P6 with the protein structure without the lid (V, Table 5). The rescored correlations were much better when all five poses were considered than when only the top-ranked conformations were considered (V, Table 5). The results of the Prime MMGBSA analysis of the top-ranked GLIDE results showed notable improvement in the correlations (V, Table 6), but not to the same degree obtained with PLANTS. Nevertheless, the correlations of all ligands were higher in GLIDE rescored with Prime MMGBSA than in PLANTS rescored with Prime MMGBSA (V, Tables 5 and 6).

Pharmacophore and 3D-QSAR models. In Phase, the pharmacophore model and atom-based 3D-QSAR model can be generated simultaneously. In the model creation, all 152 molecules were used, of which 75% and 25% were randomized to the training set and the test set, respectively. The best hypotheses were chosen based on the survival scores. From the produced 3D-QSAR models, several had $R^2 > 0.80$, $Q^2 > 0.75$, and stability values > 0.95, which indicated that the models should not be over-fitted. Based on the results of the leave-one-out test, the predictions were almost insensitive to the composition of the training set. Furthermore, the R^2 scramble values were < 0.5, indicating that the obtained correlations depend on correct data. Based on the statistics, the most promising model consisted of two hydrogen bond acceptor sites, one aromatic ring, and one hydrophobic area (V, Fig. 3). This model used three PLS components, and had $R^2 = 0.90$, $Q^2 = 0.64$, standard deviation = 0.52, F = 692.6, P = 1.937e -111, RMSE = 0.96, Pearson r = 0.87, R^2 CV = 0.8523, R^2 scramble = 0.39, and stability = 0.99.

5.6 Binding property analysis of sphingomyelin analogs (VI)

This study examined the effects of the structural properties of sphingomyelin (SM) analogs on their substrate properties with sphingomyelinase (SMase) from *Bacillus cereus*. The enzymatic activity of SM analogs was studied experimentally in micellar and monolayer substrates. Relevant to this thesis are computational studies with molecular docking and MD simulations of SMase–SM complex, which in part explain the relations between SM structures and enzyme activity.

5.6.1 Computational studies

Molecular docking and MD simulations were utilized to explore the relation of substrate structures to substrate properties. The docking of SM to the active site of SMase and the MD of the complex revealed that the 3OH group of SM coordinated with the magnesium ion and simultaneously donated a hydrogen bond to the carboxylate group of Glu53 (VI, Fig. 3). In addition, the carbonyl oxygen atom of SM accepted a hydrogen bond from Lys131, and the 2NH group formed a hydrogen bond with Asp156. Phosphorus was located in close proximity to the catalytic His296.

Because the 3OH group has a dual role in the substrate binding to SMase and the reaction center is located next to it, it was expected that 3O-methylated SM would not bind similarly as SM, and thus SMase would fail to degrade 3O-methylated SM. It was also expected that the methylation of 2NH would interfere with interactions of SM and the side chain of Asp156. The effect of the loss of induced fit on substrate binding could lead to the reduction of enzymatic activity, as observed in Asp156Gly mutation (Tamura *et al.* 1995). Experimental studies on both micellar and monolayer substrates validated the computational predictions: 3O-methylated SM was not degraded by the SMase, and 2N-methylated SM was a substrate, but it was degraded at about half the rate of SM.

In addition to the 3OH group in the long-chain base, PhytoSM has an OH-group in the C4 position. An internal hydrogen bond might form between the 4OH and 2NH or between the 4OH and carbonyl oxygen of the N-linked acyl chain. Internal hydrogen bonds might hinder the induced fit during the substrate binding, thus leading to a prolonged lag-time before the hydrolysis reaction. In computational studies, 4OH seemed not to interfere with the 3OH close to the active site, and thus did not interfere with catalysis. In experimental tests SMase readily hydrolyzed phytoPSM.

Removing two or all three terminal methyls increased the positive charge of the N, and when the positive charge becomes strong enough and the steric effect of the methyls is reduced, a new interaction could form between the N of the head group and the Glu155 and Glu250 adjacent to the active site (VI, Fig. 3). This interaction may misalign the SM molecule in the active site, especially as for Glu53 and magnesium, and thus hinder catalysis. This explains why SM lacking one methyl in the phosphocholine head group was a good substrate, but SM lacking two or three methyls failed to act as substrates for SMase.

6 DISCUSSION

6.1 Development of a novel virtual screening method

Part of this thesis is the development process of Panther method (III). Panther is a novel, simple, fast, and efficient multipurpose docking tool. In Panther, a negative image of the ligand-binding area of a protein is created by utilizing a high-quality protein crystal structure. The ligand-binding area is described as a simple atomistic shape-electrostatic model, which can be then used to screen molecular databases by using a fast similarity search algorithm. Panther was preceded by the NIB VOIDOO/FLOOD method (I, II). Although the principles of these two methods are similar, in Panther, the creation of the model is completely rebuilt.

The complementarity of the shape of the ligand and the ligand-binding site is very important in the identification of lead molecules. Previous studies on NIB VOIDOO/FLOOD demonstrated that utilizing only shape can be adequate for VS (Virtanen and Pentikäinen 2010). However, here it was shown that the addition of electrostatic information in the models generally improved the similarity searches (I, II). PDE5 (I) and ERα (II) were well suited to explore the effect of electrostatics because the ligand-binding sites have important polar interactions as well as hydrophobic interactions. It is rather self-evident that because of the polarity of the binding site, VS performed utilizing only shape of the ligand-binding area wastes part of the available, and relevant, information. On the other hand, in the automatically created Panther PR model (III), one charge point blocked the other charge point, thus restraining and misaligning the screening results. This result indicates that although building models using novel Panther algorithm is quick and relatively accurate, the positioning and utilization of the electrostatic points is not foolproof.

Moreover, the weighting of the shape and the electrostatics must be considered. In general, the AUC values were the highest in the equal weighting of the shape and the electrostatic potentials (I, II, III). However, PR was an exception because its AUC values increased significantly according to the

increased weighting of the electrostatic potential contribution in both NIB VOIDOO/FLOOD (I) and Panther screening (III). Here, basically no optimization of the parameters was done in the model creation, and individualized settings are likely to create models that describe better the electrostatic properties of the ligand-binding sites of different target proteins. The nature of the ligand-binding area, such as the ratio of hydrophobicity and polarity is likely a key factor in determining whether the addition of electrostatics in the negative image models produces negative, positive, or insignificant effects. Ultimately, the usefulness of the electrostatic component is somewhat specific to the target protein.

Protein flexibility is a crucial issue in protein structure-based VS. Its high number of degrees of freedom makes modeling the protein flexibility both computationally and physico-chemically demanding. In NIB screening, the easiest way to consider protein flexibility is by utilizing several protein crystal structures. In addition, flexibility can be incorporated into NIB screening by creating models of the snapshot structures derived from MD simulations. The importance of considering flexibility is nicely demonstrated for example by the performance of the ERα agonist and antagonist Panther models and the screening of long and narrow or short and branched molecules with glucocorticoid receptor models (III). In general, different models of the ligand-binding area are likely needed when diverse molecules and scaffold hopping are preferred as they usually are in VS.

The properties of the reference structures and the studied molecules define what can be studied with them and the kind of results that can be achieved. Thus, evaluation of goodness in any specific case will depend on the query molecule, the method, and the contents of the validation database. For example, the set of active PDE5 ligands (I) did not contain as many bulky molecules as the PDE5 inhibitor sildenafil is, which may explain the feeble results for ligand-based VS with sildenafil compared to tadalafil. Similarly, the NIB VOIDOO/FLOOD model, which closely resembled the size and the shape of sildenafil, preferred bulkier molecules, which explains the weaker results compared to tadalafil-like NIB VOIDOO/FLOOD model. The results did not indicate that sildenafil or the bulkier model could not perform better in the actual VS study, where diversity is highly valued. Thus, not only choosing the query molecule but also choosing the validation dataset has a huge effect on both the method selection and eventually the VS results. Careful attention to the composition and the utilization of benchmarking sets is essential to avoid bias (Irwin 2008, Wallach and Lilien 2011, Lagarde *et al.* 2015). Additionally, in validation, if the molecules are not experimentally tested, in theory, the presumably inactive decoy molecules may contain unidentified active compounds (as the PDE5/DUD decoys in study I or the ERα decoys in study II).

Overall, the results showed that NIB VOIDOO/FLOOD and Panther were often better than ligand-based methods or docking (I, III), and they were suitable for screening active compound in VS (II). NIB methods performed better than ligand-based methods most likely because screening with the shape

and the properties of the binding site rather than just those of a reference ligand allow more variation and thus more diverse ligands to be screened correctly. Compared to molecular docking, a state-of-the-art method in computational biology, screening utilizing NIB models is considerably faster and still relatively efficient. Overall, these results show the importance of both the shape and the electrostatics in ligand binding. In addition, they indicate that the usage of multiple protein conformations, reference ligands or negative images (NIB VOIDOO/FLOOD or Panther models) could improve the success of VS. In the further development of the Panther the strict demands are required from the method (Fig. 4).

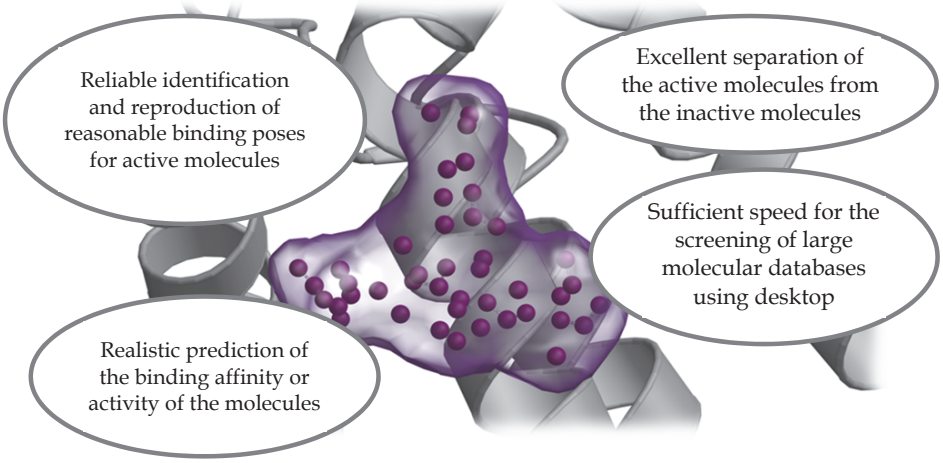


FIGURE 4 In the Panther method development the following criteria are considered.

The idea of utilizing ligand-sized cavities for drug discovery and design purposes is not new (Kleywegt and Jones 1994, Liang *et al*. 1998, Nayal and Honig 2006, Coleman and Sharp 2010, Hetényi and van der Spoel 2011). Other methods utilizing negative images in VS have been introduced (Oshiro and Kuntz 1998, Fukunishi *et al*. 2006, Ebalunode *et al*. 2008, Lee *et al*. 2009, Lee and Zhang 2012). In these methods, the ligand-binding area of the protein is modelled as a negative image, pseudoligand or virtual atoms, which may or may not carry also chemical information. The model is then used to identify potential novel compounds with methods known from ligand-based VS. The results obtained with other negative image methods also show similar trend as NIB VOIDOO/FLOOD and Panther: the efficient identification of active compounds and active molecule configurations, comparable or better AUC and enrichment ratios than other tested methods, and high diversity among the top ranking hit molecules. In general, these results support and encourage the usage of NIB methods in VS (Oshiro and Kuntz 1998, Fukunishi *et al.* 2006, Ebalunode *et al.* 2008, Lee *et al.* 2009, Lee and Zhang 2012).

6.2 Reliability of the virtual screening methods

Ligand-based and protein structure-based methods are commonly used in VS. One of the first features demanded from the VS method is the ability to separate active molecules from inactive molecules. Ligand-based screening is usually able to find active compounds; however, the requirement is that such molecules have already been established in order to use them as a reference structure. Here, the AUC values for ligand-based screening were clearly lower than the ones produced by the protein structure-based methods (I). In general, also the enrichments of the ligand-based VS were weaker than those in the protein structure-based methods. However, also studies with opposite results exist (Zhang and Muegge 2006, McGaughey *et al.* 2007, Chen *et al.* 2009, Kinnings and Jackson 2009, Krüger and Evers 2010, Mishra and Basu 2013).

Although the scoring functions of docking methods are not always efficient in the identification of active molecules with high accuracy, in some cases, they are able to find certain ligand conformations better than the ligand-based or NIB methods can (Warren *et al.* 2006, McGaughey *et al.* 2007, Kolb and Irwin 2009, Plewczynski *et al.* 2011, Grinter and Zou 2014). Therefore, molecular docking detects active ligands of certain type with a high level of certainty, which is often reflected in the relatively high early enrichment, even though the overall AUC values do not indicate successful screening results. Many previous studies suggest that molecular docking is effective in VS (Ghosh *et al*. 2006, Cavasotto *et al*. 2008, Clark 2008, Markt *et al*. 2008, Ripphausen *et al*. 2010, Murgueitio *et al*. 2012, Kumar and Zhang 2014, Cerqueira *et al*. 2015, Danishuddin and Khan 2015).

The derived pharmacophore and 3D-QSAR hypotheses were capable of separating the most and the least active SERMs (II) or phosphodiesterase 4B molecules (IV). However, although molecular diversity is taken into account in the model creation, some data of the more unusual molecules is missed because only common features and chemotypes are searched (Yang 2010). According to validation, the constructed pharmacophore and 3D-QSAR models were able to differentiate the active and inactive ligands acceptably.

Compared to ligand-based methods, screening with NIB methods allows greater freedom in ligand positioning, which enables variability in the identification of active molecules. Unlike docking, NIB methods may allow some spatial overlap between the ligand and the protein, leaving space for e.g. induced fit upon binding. This may explain the good performance of NIB methods compared to ligand-based methods and molecular docking.

In general, the post-processing of the top-ranked VS results with MD/MMGBSA improved the enrichments. For example, the post-processing of the NIB VOIDOO/FLOOD VS results produced considerably higher early enrichment than the ligand-based screening or docking (I). Interestingly, the enrichment produced by docking was diminished after the post-processing. In docking, if the sampled compounds do not match the surroundings perfectly

(e.g. some overlaps are produced or side pockets neglected), the docking algorithms usually do not rank them high, and occasionally the biologically relevant poses were ignored. However, in the post-processing stage, the protein structure is handled flexibly, and more room for molecules is offered, thereby altering their positioning and thus their scoring. Furthermore, rescoring the results using MD/MMGBSA may benefit from considering more than just one pose per ligand (V). Although post-processing with MD/MMGBSA improved the results in most cases, the results differed according to which MMGBSA method was used (I and IV vs. V).

The results showed that docking could also predict binding affinities (II, V), which is slightly surprising because it is common that using the internal scoring of docking programs seldom leads to meaningful correlations (Ferrara *et al.* 2004, Warren *et al.* 2006, McGaughey *et al.* 2007, Cross *et al.* 2009, von Korff *et al.* 2009, Feliu and Oliva 2010, Plewczynski *et al.* 2011). The comparison of the performance of the binding free energy calculation methods MMGBSA, MMPBSA, and SIE regarding their ability to identify active molecules from inactive molecules and their ability to predict experimentally determined binding affinities showed substantial differences (IV). Earlier studies have indicated that MMGBSA, MMPBSA and SIE methods could be useful and accurate tools for predicting binding affinity with high accuracy (Ferrari *et al.* 2007, Guimarães and Cardozo 2008, Rastelli *et al.* 2010, Hou *et al.* 2011a, Hou *et al.* 2011b, Mulakala and Viswanadhan 2013, Ylilauri and Pentikäinen 2013, Greenidge *et al.* 2014). Conversely, it has been shown that the results of binding free energy calculation methods can be highly case specific, depending heavily on the used parameters and the complex generation method; and moreover, variation in the performance of different IGB models may be remarkable (Ferrari *et al.* 2007, Hou *et al.* 2011a, Hou *et al.* 2011b, Greenidge *et al.* 2014).

Although the target proteins may undergo vast conformational changes upon ligand binding, earlier studies (Thompson *et al.* 2008, Rastelli *et al.* 2010) and the studies presented in this thesis suggest that the length of the MD simulation is not an important factor in binding free energy calculation (IV), and that short MD simulations or even snapshots from the beginning of the MD can produce reasonable results enabling the screening of larger number of compounds (I, IV, V). The highest AUC values and correlation coefficients for experimental activity and predicted binding affinity were often obtained by relatively short MD simulations of less than 128 ps. Additionally, the binding free energy calculation results showed that in many cases, the results originating from two different complex generation methods—molecular docking and ligand-based similarity superimposition—were quite similar (IV).

In addition to the accuracy of the method, also speed and the computational cost of the method must be taken into account. Given the size of molecular databases, a practical VS method cannot take more than a few seconds per compound. Typically, ligand-based VS methods are faster and computationally more straightforward to perform than protein structure-based methods, such as molecular docking, are. However, MD and binding free

energy calculations are even more demanding. Of the binding free energy methods investigated, MMGBSA is computationally the most efficient, while MMPBSA and SIE are more vigorous. NIB methods are protein structure-based methods, but still their computational demands are closer to ligand-based methods, which is a benefit in VS. For example, NIB VOIDOO/FLOOD and Panther are not strenuous if their computational costs are compared with ligand-based screening or molecular docking (I, II, III).

Many biases can result from comparing VS methods, such as the selection of targets, active molecules and decoys, adjustable parameters, means used to evaluate of the performance of the methods, and so forth. For example, comparisons within and between AUCs and enrichment values are challenging and may be potentially biased (Truchon and Bayly 2007, Good and Oprea 2008, Jain 2008, Jain and Nicholls 2008, Sheridan 2008). The methods used for enrichment calculation vary greatly, and the percentages selected for comparisons differ. The preparation of the validation molecular databases greatly affects the molecules, so that the seemingly same dataset may yield different results and thus falsify the comparisons. In addition, the researcher is a potential source of subjective bias. For example, although visual inspection of the molecules is not objective, it is important to understand how a compound fits the ligand-binding area of the protein, aligns with a template molecule, or is mapped onto a pharmacophore, and in such cases, the subjective position of the researcher may bias the decision making.

7 CONCLUSIONS

The variety of VS methods is enormous in terms of their application area, working principles, and usability. All methods have both advantages and disadvantages. Ligand-based methods are extremely fast in VS. The pharmacophore and 3D-QSAR methods find the most active ligands, but their prediction capacity is highly chemotype dependent. Docking can be used to identify active molecules and to estimate binding affinities. NIB screening is efficient, and the addition of electrostatic information to the models improved it in most cases. Considering the protein flexibility in the form of multiple protein crystal structures or snapshots from MD brought additional value to the search for some targets. NIB screening, such as the novel Panther method, can be especially useful when there are no known active ligands, the number of known actives is low, or when docking is either ineffective or too slow. Additionally, the post-processing of the initially ranked molecules using the MD/MMGBSA produced significant early enrichment in the recognition of active ligands and thus the additional usage of MMGBSA analysis could serve as a tool to identify the most promising compounds. However, the findings also showed differences in the performance of the binding free energy calculation methods in active ligand recognition and in their ability to predict the activity of the compounds. According the results, complexes for binding free energy calculations can be generated by using either the docking or the ligand-based similarity method. Regarding the length of the MD simulation, the results suggest that also short simulation or even a single energy-minimized snapshot may be sufficient for the efficient binding free energy calculations. However, for the prediction of actual binding affinities, a longer MD simulation may be required.

The comparison of the use of the above-mentioned methods in VS showed their differing abilities to identify active molecules and predict binding affinities, and moreover could also further the understanding of the advantages and limitations of these methods. However, it is difficult to recommend any method over other methods. The results of this study align well with earlier VS studies suggesting that computational methods could be used successfully to identify novel ligands, especially when multiple methods are combined. However,

caution and meticulousness are needed because the performance of computational methods may be case specific. Therefore, prior to the VS study, the careful validation of methods is recommended. Ultimately, the features of the computational method define its applicability and therefore also the quality of the results. Moreover, the reference structures, validation molecules, and studied compounds determine for their part what can be discovered. Thus, not only the methodology but also the molecules used must be considered prudently.

Acknowledgements

This work was carried out during 2011-2015 at the University of Jyväskylä, at the Department of Biological and Environmental Science and Nanoscience Center. The thesis work was funded by the National Doctoral Programme in Nanoscience (NGS-NANO). The computational resources were provided by CSC – IT Center for Science.

First of all I would like to thank my supervisor Dr. Olli Pentikäinen. We have had many lively and fruitful discussions which sometimes move things forward and sometimes on the other hand… Anyway, it is nice to be able to say frankly what one really thinks.

Next I would like to thank my opponent Dr. Daniela Schuster. You know, Finland in November… Maybe not the most tempting travel destination but still you accepted the invitation and came.

Then I acknowledge the reviewers of the thesis Dr. Tuomo Laitinen and Dr. Adyary Fallarero and my follow-up group members Dr. Maija Lahtela-Kakkonen and Prof. Hannu Häkkinen. You all have given me valuable comments and advice from different angles at different stages of this project.

Pekka Postila, you were one of my first contacts to the real science world. Thank you for that. Salla Virtanen, our projects have walked nicely hand-in-hand. If we had spent more time in the same office who knows what would have happened. Mikko Ylilauri, there are many things to say to you. In a condensed form, I think we are more than just former co-workers, we are friends. Sanna Rauhamäki, you are Sanna and usually that is enough. Mira Kuusisto and Kari Salokas, you both are bright young scientists and fun company outside office as well. I would also like to thank my other colleagues, especially former CBL members Elisa Nurminen and Jarkko Koivunen.

Jonne, during this process you were my closest peer in the office and a good friend in the outside world. Thanks for tea and sympathy and that you had guts to climb all the way up to the Ivory tower frequently.

I am grateful to my friends outside work environment. You take my mind of the academic business in numerous ways which is incredibly important. Here a special mention goes to Kati, you are truly one of a kind and our history goes way beyond. To me, nothing can ever erase that. I also want to thank my parents and my little brother, you have been supportive.

Last but not least, Sakari. You are always there when I need you most. You really made this thesis possible in every aspect. In fact, you make many impossible things possible. Thank you for your support, help, patience, love, care… There are no words to describe my gratitude.

YHTEENVETO (RÉSUMÉ IN FINNISH)

Laskennallisia tutkimuksia biomolekyylien seulonnasta ja vuorovaikutuksista

Lääkkeet ovat kemiallisia yhdisteitä, joilla hoidetaan tai ennaltaehkäistään sairauksia. Tavallisesti lääkkeiden vaikuttavat aineet ovat pieniä orgaanisia molekyylejä. Lääkkeiden sisältämät aktiiviset pienmolekyylit vaikuttavat elimistössä muokkaamalla proteiinien toimintaa solujen signalointireiteillä ja metaboliassa. Proteiinit ovat suuria biologisia makromolekyylejä, jotka toimivat käytännöllisesti katsoen kaikissa solun prosesseissa. Esimerkiksi entsyymit katalysoivat kemiallisia reaktioita ja reseptoriproteiinit toimivat solun viestiliikenteessä. Koska proteiinit ovat hyvin monimuotoisia, myös niihin sitoutuvat lääkeaineet ovat erilaisia ja vaikuttavat kohdeproteiineihinsa monella tavalla, tyypillisesti joko vahvistamalla tai heikentämällä proteiinin toimintaa.

Uuden lääkkeen saattaminen markkinoille on pitkä ja kallis prosessi. Tyypillisesti lääkekehitysprojekti koostuu kahdesta vaiheesta: uusien lääkeaihioiden etsinnästä ja niiden kehittämisestä. Etsintävaiheessa satojen tuhansien molekyylien aktiivisuus voidaan seuloa kokeellisesti, mutta vain muutama tutkituista yhdisteistä on riittävän lupaava edetäkseen kehitysvaiheeseen. Kehitysvaiheessa yhdisteiden turvallisuus ja tehokkuus arvioidaan tarkemmin, yhdisteillä tehdään kliinisiä kokeilta ja niitä muokataan parempien ominaisuuksien saavuttamiseksi. Lopulta vain hyvin pieni osa tutkituista molekyyleistä hyväksytään uusiksi lääkeaineiksi. Tämän tehottomuuden vuoksi rationaalinen lääkeainesuunnittelu tarvitsee uusia menetelmiä, jotka sekä nopeuttavat lääkeaihiomolekyylien etsintää että tunnistavat aktiiviset aihiomolekyylit tarkemmin.

Virtuaaliseulonta on tehokas tapa suodattaa tietokoneavusteisesti suuria molekyylitietokantoja lääkeaihiomolekyylien löytämiseksi. Tämä mahdollistaa kokeellisissa tutkimuksissa keskittymisen vain oletettavasti parhaimpien molekyylien ominaisuuksien tarkasteluun tehostaen näin lääkekehitysprosessia. Virtuaaliseulontamenetelmä voidaan suunnitella hyödyntämään niin tunnettujen sitoutuvien molekyylien eli ligandien kuin kohdeproteiinin rakenteellista informaatiota. Virtuaaliseulontamenetelmien avulla voidaan molekyylien etsimisen lisäksi ennustaa molekyylien ominaisuuksia kuten molekyylien myrkyllisyyttä ja metaboloitumista elimistössä.

Tämän väitöskirjan osatöissä jatkettiin uuden virtuaaliseulontamenetelmän kehittämistä. Tämä menetelmä hyödyntää kohdeproteiinin ligandinsitomisalueen rakenteellista ja kemiallista tietoa molekyylitietokantojen nopeaan seulontaan. Lisäksi väitöskirjan osatöissä testattiin ja vertailtiin useiden laskennallisten menetelmien käyttöä virtuaaliseulonnassa ja molekyylien sitoutumisen ennustamisessa.

Olemassa olevien virtuaaliseulontamenetelmien kirjo on valtava. Menetelmät eroavat toisistaan toimintaperiaatteidensa ja käyttötarkoituksensa osalta. Kaikissa käytössä olevissa menetelmissä on hyvät ja huonot puolensa. Ligandipohjaiset eli tunnettuja sitoutuvia molekyylejä hyödyntävät menetelmät ovat

erittäin nopeita virtuaaliseulontatyökaluja, mutta toisaalta hyvin riippuvaisia malleina käytetyistä molekyyleistä. Farmakoforimallit ja kolmiulotteinen rakenne-aktiivisuusanalyysi tunnistavat aktiiviset molekyylit tarkasti, mutta niiden kyky ennustaa molekyylejä riippuu niin ikään mallin rakentamisessa käytetyistä molekyyleistä. Telakoinnissa molekyylejä sovitetaan kohdeproteiinin ligandinsitomisalueelle, jolloin voidaan arvioida parhaiten kohteeseen soveltuvat molekyylit sekä niiden sitoutumisvoimakkuus. Kohdeproteiinin ligandinsitomisalueesta luotuun negatiivikuvaan perustuva seulonta on tehokasta sekä laskennallisesti että molekyylien tunnistamisessa.

Tämän väitöskirjan osatöissä esitelty ligandinsitomisalueen negatiivikuvaan perustuva virtuaaliseulontamenetelmä on erityisen hyödyllinen kun tunnettuja sitoutuvia ligandeja ei ole tarpeeksi tai kun molekyylien telakointi on joko tehotonta tai liian aikaavievää. Lisäksi todettiin, että sähköstaattisen informaation lisääminen negatiivikuvamalliin tehostaa molekyylien etsintää. Joillekin kohdeproteiineille proteiinirakenteiden joustavuuden huomiointi esimerkiksi käyttämällä useita kokeellisesti ratkaistuja proteiinikiderakenteita tai molekyylidynamiikkasimulaatiosta poimittuja rakenteita niin ikään parantaa virtuaaliseulonnan tulosta. Lisäksi nopeilla virtuaaliseulontamenetelmillä parhaiksi arvioitujen molekyylien pisteyttäminen uudelleen tarkemmalla ja laskennallisesti vaativammalla molekyylien sitoutumisenergiaa määrittävillä menetelmällä osoittautui hyödylliseksi, koska se tehosti edelleen aktiivisimpien molekyylien erottamista heikommista sitoutujista. Tutkimuksessa havaittiin myös, että molekyylien sitoutumisenergiaa määrittävien menetelmien välillä oli eroja sekä aktiivisten molekyylien tunnistamisessa että kyvyssä ennustaa molekyylien aktiivisuus. Tulokset kuitenkin osoittivat, että sitoutumisenergiaa määrittäviin kokeisiin proteiini-ligandikompleksit voidaan luoda kummalla tahansa menetelmällä joko ligandipohjaisella samankaltaisuusvertailulla tai telakoimalla. Sitoutumisenergian määrittämistä edeltävien molekyylidynamiikkasimulaatioiden pituudeksi voi joissain tapauksissa riittää joko hyvin lyhyt simulaatio tai jopa pelkkä proteiini-ligandikompleksin energian minimointi. On kuitenkin mahdollista, että todellisten sitoutumisvoimakkuuksien arviointiin vaaditaan pidempiä molekyylidynamiikkasimulaatioita.

Tämän väitöskirjan osatöissä suoritetut virtuaaliseulontamenetelmien vertailut osoittavat eroja menetelmien kyvyssä erotella aktiivisia ligandeja inaktiivisista molekyyleistä ja kyvyssä ennustaa molekyylien sitoutumisen voimakkuutta. Vertailut myös auttavat hahmottamaan testattujen menetelmien vahvuuksia ja heikkouksia. Tutkimuksen tulokset ovat samankaltaisia aiempien tutkimusten kanssa ja niiden perusteella voidaan suositella laskennallisten menetelmien käyttöä uusien molekyylien etsintään, mutta yhden menetelmän käyttöä ylitse muiden on vaikeata suositella. Koska tulokset voivat olla jossain määrin tapauskohtaisia, menetelmien soveltuvuuden varmentaminen ennen suurta virtuaaliseulontatutkimusta on tärkeää ja oletettavasti parhaaseen tulokseen päästään kun käytetään useita menetelmiä harkitusti yhtä aikaa.

REFERENCES

Åqvist J. 1990. Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* 94: 8021–8024.

Armstrong M.S., Finn P.W., Morris G.M. & Richards W.G. 2011. Improving the accuracy of ultrafast ligand-based screening: Incorporating lipophilicity into ElectroShape as an extra dimension. *J. Comput. Aided. Mol. Des.* 25: 785–790.

Armstrong M.S., Morris G.M., Finn P.W., Sharma R., Moretti L., Cooper R.I. & Richards W.G. 2010. ElectroShape: Fast molecular similarity calculations incorporating shape, chirality and electrostatics. *J. Comput. Aided. Mol. Des.* 24: 789–801.

Baell J. & Walters M.A. 2014. Chemical con artists foil drug discovery. *Nature* 513: 481–483.

Baell J.B. & Holloway G. a. 2010. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* 53: 2719–2740.

Bajusz D., Rácz A. & Héberger K. 2015. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* 7: 20.

Ballester P.J. 2011. Ultrafast shape recognition: method and applications. *Future Med. Chem.* 3: 65–78.

Ballester P.J., Finn P.W. & Richards W.G. 2009. Ultrafast shape recognition: Evaluating a new ligand-based virtual screening technology. *J. Mol. Graph. Model.* 27: 836–845.

Ballester P.J. & Richards W.G. 2007. Ultrafast shape recognition to search compound databases for similar molecular shapes. *J. Comput. Chem.* 28: 1711–1723.

Barnard J.M. & Downs G.M. 1997. Chemical fragment generation and clustering software. *J. Chem. Inf. Model.* 37: 141–142.

Bellis L.J., Akhtar R., Al-Lazikani B., Atkinson F., Bento A.P., Chambers J., Davies M., Gaulton A., Hersey A., Ikeda K., Krüger F.A., Light Y., McGlinchey S., Santos R., Stauch B. & Overington J.P. 2011. Collation and data-mining of literature bioactivity data for drug discovery. *Biochem. Soc. Trans.* 39: 1365–1370.

Bender A., Mussa H.Y., Glen R.C. & Reiling S. 2004. Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): Evaluation of performance. *J. Chem. Inf. Comput. Sci.* 44: 1708–1718.

Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., DiNola A. & Haak J.R. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81: 3684–3690.

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

Boström J. 2001. Reproducing the conformations of protein-bound ligands: A critical evaluation of several popular conformational searching tools. *J. Comput. Aided. Mol. Des.* 15: 1137–1152.

Cannon E.O., Nigsch F. & Mitchell J.B.O. 2008. A novel hybrid ultrafast shape descriptor method for use in virtual screening. *Chem. Cent. J.* 2: 3.

Case D.A., Darden T.A., Cheatham T.E.I., Simmerling C.L., Wang J., Duke R.E., Luo R., Crowley M., Walker R.C., Zhang W., Merz K.M., Wang B., Hayik S., Roitberg A., Seabra G., Kolossváry I., Wong K.F., Paesani F., Vanicek J., Wu X., Brozell S.R., Steinbrecher T., Gohlke H., Yang L., Tan C., Mongan J., Hornak V., Cui G., Mathews D.H., Seetin M.G., Sagui C., Babin V. & Kollman P.A. 2008. AMBER 10. University of California, San Francisco.

Cavasotto C.N., Orry A.J.W., Murgolo N.J., Czarniecki M.F., Kocsi S.A., Hawes B.E., O'Neill K.A., Hine H., Burton M.S., Voigt J.H., Abagyan R.A., Bayne M.L. & Monsma F.J. 2008. Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. *J. Med. Chem.* 51: 581–588.

Cerqueira N.M.F.S.A., Gesto D., Oliveira E.F., Santos-Martins D., Brás N.F., Sousa S.F., Fernandez P.A. & Ramos M.J. 2015. Receptor-based virtual screening protocol for drug discovery. *Arch. Biochem. Biophys.* 582: 56–67.

Chen Y.-C. 2015. Beware of docking! *Trends Pharmacol. Sci.* 36: 78–95.

Chen Z., Li H.L., Zhang Q.J., Bao X.G., Yu K.Q., Luo X.M., Zhu W.L. & Jiang H.L. 2009. Pharmacophore-based virtual screening versus docking-based virtual screening: a benchmark comparison against eight targets. *Acta Pharmacol. Sin.* 30: 1694–1708.

Cheng T., Li Q., Zhou Z., Wang Y. & Bryant S.H. 2012. Structure-Based Virtual Screening for Drug Discovery: a Problem-Centric Review. *AAPS J.* 14: 133–141.

Cherkasov A., Muratov E.N., Fourches D., Varnek A., Baskin I.I., Cronin M., Dearden J., Gramatica P., Martin Y.C., Todeschini R., Consonni V., Kuz'Min V.E., Cramer R., Benigni R., Yang C., Rathman J., Terfloth L., Gasteiger J., Richard A. & Tropsha A. 2014. QSAR modeling: Where have you been? Where are you going to? *J. Med. Chem.* 57: 4977–5010.

Clark D.E. 2008. What has virtual screening ever done for drug discovery? *Expert Opin. Drug Discov.* 3: 841–851.

Cleves A.E. & Jain A.N. 2015. Knowledge-guided docking: accurate prospective prediction of bound configurations of novel ligands using Surflex-Dock. *J. Comput. Aided. Mol. Des.* 29: 485–509.

Coleman R.G. & Sharp K. a. 2010. Protein pockets: Inventory, shape, and comparison. *J. Chem. Inf. Model.* 50: 589–603.

Congreve M., Carr R., Murray C. & Jhoti H. 2003. A "Rule of Three" for fragment-based lead discovery? *Drug Discov. Today* 8: 876–877.

Connolly M.L. 1983. Analytical molecular surface calculation. *J. Appl. Crystallogr.* 16: 548–558.

Cortés-Cabrera A., Morris G.M., Finn P.W., Morreale A. & Gago F. 2013. Comparison of ultra-fast 2D and 3D ligand and target descriptors for side effect prediction and network analysis in polypharmacology. *Br. J. Pharmacol.* 170: 557–567.

Cramer R.D., Patterson D.E. & Bunce J.D. 1988. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* 110: 5959–5967.

Cross J.B., Thompson D.C., Rai B.K., Baber J.C., Fan K.Y., Hu Y. & Humblet C. 2009. Comparison of several molecular docking programs: Pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* 49: 1455–1474.

Cui Q., Sulea T., Schrag J.D., Munger C., Hung M.N., Naïm M., Cygler M. & Purisima E.O. 2008. Molecular dynamics-solvated interaction energy studies of protein-protein interactions: The MP1-p14 Scaffolding Complex. *J. Mol. Biol.* 379: 787–802.

Dahlin J., Nissink W., Strasser J.M., Francis S., Higgins L., Zhou H., Zhang Z. & Walters M.A. 2015. PAINS in the assay: chemical mechanisms of assay interference and promiscuous enzymatic inhibition observed during a sulfhydryl-scavenging HTS. *J. Med. Chem.* 58: 2091–2113.

Danishuddin M. & Khan A.U. 2015. Structure based virtual screening to discover putative drug candidates: Necessary considerations and successful case studies. *Methods* 71: 135–145.

Darden T., York D. & Pedersen L. 1993. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98: 10089.

Deng Z., Chuaqui C. & Singh J. 2004. Structural Interaction Fingerprint (SIFt): A novel method for analyzing three-dimensional protein-ligand binding interactions. *J. Med. Chem.* 47: 337–344.

Dixon S.L., Smondyrev A.M., Knoll E.H., Rao S.N., Shaw D.E. & Friesner R.A. 2006a. PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J. Comput. Aided. Mol. Des.* 20: 647–671.

Dixon S.L., Smondyrev A.M. & Rao S.N. 2006b. PHASE: A novel approach to pharmacophore modeling and 3D database searching. *Chem. Biol. Drug Des.* 67: 370–372.

Dobi K., Hajdú I., Flachner B., Fabó G., Szaszkó M., Bognár M., Magyar C., Simon I., Szisz D., Lõrincz Z., Cseh S. & Dormán G. 2014. Combination of 2D/3D ligand-based similarity search in rapid virtual screening from multimillion compound repositories. Selection and biological evaluation of potential PDE4 and PDE5 inhibitors. *Molecules* 19: 7008–7039.

Dong F., Olsen B. & Baker N.A. 2008. Computational methods for biomolecular electrostatics. *Methods Cell Biol.* 84: 843–870.

Drwal M.N. & Griffith R. 2013. Combination of ligand- and structure-based methods in virtual screening. *Drug Discov. Today Technol.* 10: .

Duan J., Dixon S.L., Lowrie J.F. & Sherman W. 2010. Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. *J. Mol. Graph. Model.* 29: 157–170.

Duan Y., Wu C., Chowdhury S., Lee M.C., Xiong G., Zhang W., Yang R., Cieplak P., Luo R., Lee T., Caldwell J., Wang J. & Kollman P. 2003. A point-charge force field for molecular mechanics simulations of proteins based on

condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24: 1999–2012.

Du-Cuny L., Chen L. & Zhang S. 2011. A critical assessment of combined ligand- and structure-based approaches to hERG channel blocker modeling. *J. Chem. Inf. Model.* 51: 2948–2960.

Ebalunode J.O., Ouyang Z., Liang J. & Zheng W. 2008. Novel approach to structure-based pharmacophore search using computational geometry and shape matching techniques. *J. Chem. Inf. Model.* 48: 889–901.

Evers A., Hessler G., Matter H. & Klabunde T. 2005. Virtual screening of biogenic amine-binding G-protein coupled receptors: Comparative evaluation of protein- and ligand-based virtual screening protocols. *J. Med. Chem.* 48: 5448–5465.

Feliu E. & Oliva B. 2010. How different from random are docking predictions when ranked by scoring functions? *Proteins Struct. Funct. Bioinforma.* 78: 3376–3385.

Ferrara P., Gohlke H., Price D.J., Klebe G., Iii C.L.B. & Brooks 3rd C.L. 2004. Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.* 47: 3032–3047.

Ferrari A.M., Degliesposti G., Sgobba M. & Rastelli G. 2007. Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. *Bioorganic Med. Chem.* 15: 7865–7877.

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. 2004. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47: 1739–1749.

Friesner R.A., Murphy R.B., Reparsky M.P., Frye L.L., Greenwood J.R., Halgren T.A., Sanschagrin P.C. & Mainz D.T. 2006. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* 49: 6177–6196.

Fukunishi Y., Kubota S., Kanai C. & Nakamura H. 2006. A virtual active compound produced from the negative image of a ligand-binding pocket, and its application to in-silico drug screening. *J. Comput. Aided. Mol. Des.* 20: 237–248.

Genheden S. & Ryde U. 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* 10: 449–461.

Ghosh S., Nie A., An J. & Huang Z. 2006. Structure-based virtual screening of chemical libraries for drug discovery. *Curr. Opin. Chem. Biol.* 10: 194–202.

Golbraikh A., Shen M., Xiao Z., Xiao Y.-D., Lee K.-H. & Tropsha A. 2003. Rational selection of training and test sets for the development of validated QSAR models. *J. Comput. Aided. Mol. Des.* 17: 241–53.

Good A.C. & Oprea T.I. 2008. Optimization of CAMD techniques 3. Virtual screening enrichment studies: A help or hindrance in tool selection? *J. Comput. Aided. Mol. Des.* 22: 169–178.

Goodsell D.S., Morris G.M. & Olson A.J. 1996. Automated docking of flexible ligands: Applications of AutoDock. *J. Mol. Recognit.* 9: 1–5.

Greenidge P.A., Kramer C., Mozziconacci J.C. & Sherman W. 2014. Improving docking results via reranking of ensembles of ligand poses in multiple X-ray protein conformations with MM-GBSA. *J. Chem. Inf. Model.* 54: 2697–2717.

Grinter S.Z. & Zou X. 2014. Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design. *Molecules* 19: 10150–10176.

Guimarães C.R.W. & Cardozo M. 2008. MM-GB/SA rescoring of docking poses in structure-based lead optimization. *J. Chem. Inf. Model.* 48: 958–70.

Gund P. 1979. Pharmacophoric pattern searching and receptor mapping. *Ann. Reports Med. Chem.* 14: 299–308.

Güner O.F. 2002. History and evolution of the pharmacophore concept in computer-aided drug design. *Curr. Top. Med. Chem.* 2: 1321–1332.

Güner O.F. & Bowen J.P. 2013. Pharmacophore modeling for ADME. *Curr. Top. Med. Chem.* 13: 1327–1342.

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

Halgren T.A. 1996. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* 17: 490–519.

Halperin I., Ma B., Wolfson H. & Nussinov R. 2002. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins Struct. Funct. Genet.* 47: 409–443.

Hanley J.A. & McNeil B.J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 29–36.

Hansch C. 1969. A Quantitative Approach to Biochemical Structure-Activity Relationships. *Chem. Res.* 2: 232–239.

Hansch C., Deutsch E.W. & Smith R.N. 1965. The use of substituent constants and regression analysis in the study of enzymatic reaction mechanisms. *J. Am. Chem. Soc.* 87: 2738–2742.

Hansch C. & Fujita T. 1964. ρ-σ-π Analysis. A Method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* 86: 1616–1626.

Heikamp K. & Bajorath J. 2013. The future of virtual compound screening. *Chem. Biol. Drug Des.* 81: 33–40.

Hetényi C. & van der Spoel D. 2011. Toward prediction of functional protein pockets using blind docking and pocket search algorithms. *Protein Sci.* 20: 880–893.

Hevener K.E., Mehboob S., Su P.C., Truong K., Boci T., Deng J., Ghassemi M., Cook J.L. & Johnson M.E. 2012. Discovery of a novel and potent class of F. tularensis enoyl-reductase (FabI) inhibitors by molecular shape and electrostatic matching. *J. Med. Chem.* 55: 268–279.

Hou T., Wang J., Li Y. & Wang W. 2011a. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy

calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* 51: 69–82.

Hou T., Wang J., Li Y. & Wang W. 2011b. Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. the accuracy of ranking poses generated from docking. *J. Comput. Chem.* 32: 866–877.

Hu G., Kuang G., Xiao W., Li W., Liu G. & Tang Y. 2012. Performance evaluation of 2D fingerprint and 3D shape similarity methods in virtual screening. *J. Chem. Inf. Model.* 52: 1103–1113.

Hu Y. & Bajorath J. 2014. Influence of search parameters and criteria on compound selection, promiscuity, and pan assay interference characteristics. *J. Chem. Inf. Model.* 54: 3056–3066.

Huang N., Shoichet B.K. & Irwin J.J. 2006. Benchmarking sets for molecular docking. *J. Med. Chem.* 49: 6789–6801.

Huang S.Y. & Zou X. 2010. Advances and challenges in protein-ligand docking. *Int. J. Mol. Sci.* 11: 3016–3034.

Irwin J.J. 2008. Community benchmarks for virtual screening. *J. Comput. Aided. Mol. Des.* 22: 193–199.

Irwin J.J. & Shoichet B.K. 2005. ZINC - A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* 45: 177–182.

Jain A.N. 2008. Bias, reporting, and sharing: Computational evaluations of docking methods. *J. Comput. Aided. Mol. Des.* 22: 201–212.

Jain A.N. 2003. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* 46: 499–511.

Jain A.N. & Nicholls A. 2008. Recommendations for evaluation of computational methods. *J. Comput. Aided. Mol. Des.* 22: 133–139.

Jakalian A., Bush B.L., Jack D.B. & Bayly C.I. 2000. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* 21: 132–146.

Johnson M., Lajiness M. & Maggiora G. 1989. Molecular similarity: a basis for designing drug screening programs. *Prog. Clin. Biol. Res.* 291: 176–71.

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

Kinnings S.L. & Jackson R.M. 2009. LigMatch: A multiple structure-based ligand matching method for 3D virtual screening. *J. Chem. Inf. Model.* 49: 2056–2066.

Kirchmair J., Distinto S., Markt P., Schuster D., Spitzer G.M., Liedl K.R. & Wolber G. 2009. How to optimize shape-based virtual screening: choosing the right query and including chemical information. *J. Chem. Inf. Model.* 49: 678–692.

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

Kleywegt G.J. & Jones T.A. 1994. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 50: 178–185.

Kolb P. & Irwin J.J. 2009. Docking screens: right for the right reasons? *Curr. Top. Med. Chem.* 9: 755–770.

Kollman P. 1993. Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* 93: 2395–2417.

Kollman P. a., Massova I., Reyes C., Kuhn B., Huo S., Chong L., Lee M., Lee T., Duan Y., Wang W., Donini O., Cieplak P., Srinivasan J., Case D. a. & Cheatham T.E. 2000. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33: 889–897.

Korb O., Stützle T. & Exner T.E. 2009. Empirical scoring functions for advanced Protein-Ligand docking with PLANTS. *J. Chem. Inf. Model.* 49: 84–96.

von Korff M., Freyss J. & Sander T. 2009. Comparison of ligand- and structure-based virtual screening on the DUD data set. *J. Chem. Inf. Model.* 49: 209–231.

Kortagere S., Krasowski M.D. & Ekins S. 2009. The importance of discerning shape in molecular pharmacology. *Trends Pharmacol. Sci.* 30: 138–147.

Kramer B., Rarey M. & Lengauer T. 1999. Evaluation of the FlexX incremental construction algorithm for protein- ligand docking. *Proteins Struct. Funct. Genet.* 37: 228–241.

Krammer A., Kirchhoff P.D., Jiang X., Venkatachalam C.M. & Waldman M. 2005. LigScore: A novel scoring function for predicting binding affinities. *J. Mol. Graph. Model.* 23: 395–407.

Kraulis P.J. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24: 946–950.

Krüger D.M. & Evers A. 2010. Comparison of structure- and ligand-based virtual screening protocols considering hit list complementarity and enrichment factors. *ChemMedChem* 5: 148–158.

Kubinyi H. 1997. QSAR and 3D QSAR in drug design. Part 1: Methodology. *Drug Discov. Today* 2: 457–467.

Kumar A. & Zhang K.Y.J. 2014. Hierarchical virtual screening approaches in small molecule drug discovery. *Methods*.

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

Lagarde N., Zagury J.-F. & Montes M. 2015. Benchmarking data sets for the evaluation of virtual ligand screening methods: Review and perspectives. *J. Chem. Inf. Model.* 55: 1297–1307.

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. 2009. DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA* 15: 1219–1230.

Lavecchia A. & Di Giovanni C. 2013. Virtual screening strategies in drug discovery: a critical review. *Curr. Med. Chem.* 20: 2839–60.

Leach A.R., Gillet V.J., Lewis R.A. & Taylor R. 2010. Three-dimensional pharmacophore methods in drug discovery. *J. Med. Chem.* 53: 539–558.

Lee H.S., Lee C.S., Kim J.S., Kim D.H. & Choe H. 2009. Improving virtual screening performance against conformational variations of receptors by shape matching with ligand binding pocket. *J. Chem. Inf. Model.* 49: 2419–2428.

Lee H.S. & Zhang Y. 2012. BSP-SLIM: A blind low-resolution ligand-protein docking approach using predicted protein structures. *Proteins Struct. Funct. Bioinforma.* 80: 93–110.

Lehtonen J. V, Still D.J., Rantanen V. V, Ekholm J., Bjorkland D., Iftikhar Z., Huhtala M., Repo S., Jussila A., Jaakkola J., Pentikainen O., Nyronen T., Salminen T., Gyllenberg M. & Johnson M.S. 2004. BODIL: a molecular modeling environment for structure-function analysis and drug design. *J. Comput. Aided. Mol. Des.* 18: 401–419.

Li J., Abel R., Zhu K., Cao Y., Zhao S. & Friesner R.A. 2011. The VSGB 2.0 model: A next generation energy model for high resolution protein structure modeling. *Proteins* 79: 2794–2812.

Liang J., Edelsbrunner H. & Woodward C. 1998. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* 7: 1884–1897.

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

Lill M.A. 2011. Efficient incorporation of protein flexibility and dynamics into molecular docking simulations. *Biochemistry* 50: 6157–6169.

Lipinski C.A., Lombardo F., Dominy B.W. & Feeney P.J. 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23: 3–25.

Lybrand T.P., McCammon J.A. & Wipff G. 1986. Theoretical calculation of relative binding affinity in host-guest systems. *Proc. Natl. Acad. Sci. U. S. A.* 83: 833–835.

Maggiora G., Vogt M., Stumpfe D. & Bajorath J. 2014. Molecular similarity in medicinal chemistry. *J. Med. Chem.* 57: 3186–3204.

Markt P., Petersen R.K., Flindt E.N., Kristiansen K., Kirchmair J., Spitzer G., Distinto S., Schuster D., Wolber G., Laggner C. & Langer T. 2008. Discovery of novel PPAR ligands by a virtual screening approach based on pharmacophore modeling, 3D shape, and electrostatic similarity screening. *J. Med. Chem.* 51: 6303–6317.

Martin Y.C., Jarboe C.H., Krause R.A., Lynn K.R., Dunnigan D. & Holland J.B. 1973. Potential anti-parkinson drugs designed by receptor mapping. *J. Med. Chem.* 16: 147–150.

Massova I. & Kollman P.A. 2000. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discov. Des.* 18: 113–135.

McGann M. 2011. FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* 51: 578–596.

McGann M.R., Almond H.R., Nicholls A., Grant J.A. & Brown F.K. 2003. Gaussian docking functions. *Biopolymers* 68: 76–90.

McGaughey G.B., Sheridan R.P., Bayly C.I., Culberson J.C., Kreatsoulas C., Lindsley S., Maiorov V., Truchon J.F. & Cornell W.D. 2007. Comparison of

topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* 47: 1504–1519.

Merritt E.A. & Bacon D.J. 1997. Raster3D: Photorealistic Molecular Graphics. *Methods Enzymol.* 277: 505–524.

Metz C.E. 1978. Basic principles of ROC analysis. *Semin. Nucl. Med.* 8: 283–298.

Mishra N. & Basu A. 2013. Exploring different virtual screening strategies for acetylcholinesterase inhibitors. *Biomed Res. Int.* 2013: .

Mochalkin I., Miller J.R., Narasimhan L., Thanabal V., Erdman P., Cox P.B., Prasad J.V.N.V., Lightle S., Huband M.D. & Stover C.K. 2009. Discovery of antibacterial biotin carboxylase inhibitors by virtual screening and fragment-based approaches. *ACS Chem. Biol.* 4: 473–483.

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

Mulakala C. & Viswanadhan V.N. 2013. Could MM-GBSA be accurate enough for calculation of absolute protein/ligand binding free energies? *J. Mol. Graph. Model.* 46: 41–51.

Murgueitio M.S., Bermudez M., Mortier J. & Wolber G. 2012. In silico virtual screening approaches for anti-viral drug discovery. *Drug Discov. Today Technol.* 9: .

Mysinger M.M., Carchia M., Irwin J.J. & Shoichet B.K. 2012. Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *J. Med. Chem.* 55: 6582–6594.

Naïm M., Bhat S., Rankin K.N., Dennis S., Chowdhury S.F., Siddiqi I., Drabik P., Sulea T., Bayly C.I., Jakalian A. & Purisima E.O. 2007. Solvated Interaction Energy (SIE) for scoring protein-ligand binding affinities. 1. Exploring the parameter space. *J. Chem. Inf. Model.* 47: 122–133.

Nayal M. & Honig B. 2006. On the nature of cavities on protein surfaces: Application to the identification of drug-binding sites. *Proteins Struct. Funct. Genet.* 63: 892–906.

Nicholls A., McGaughey G.B., Sheridan R.P., Good A.C., Warren G., Mathieu M., Muchmore S.W., Brown S.P., Grant J.A., Haigh J. a., Nevins N., Jain A.N. & Kelley B. 2010. Molecular shape and medicinal chemistry: a perspective. *J. Med. Chem.* 53: 3862–3886.

Niinivehmas S.P., Virtanen S.I., Lehtonen J. V., Postila P.A. & Pentikäinen O.T. 2011. Comparison of virtual high-throughput screening methods for the identification of phosphodiesterase-5 inhibitors. *J. Chem. Inf. Model.* 51: 1353–1363.

Noha S., Fischer K., Koeberle A., Garscha U., Werz O. & Schuster D. 2015. Discovery of novel, non-acidic mPGES-1 inhibitors by virtual screening with a multistep protocol. *Bioorg. Med. Chem.* 23: 4839–4845.

O'Boyle N.M., Banck M., James C.A., Morley C., Vandermeersch T. & Hutchison G.R. 2011. Open Babel: An open chemical toolbox. *J. Cheminform.* 3: 33.

Onufriev A., Bashford D. & Case D.A. 2004. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins Struct. Funct. Genet.* 55: 383–394.

Oshiro C.M. & Kuntz I.D. 1998. Characterization of receptors with a new negative image: Use in molecular docking and lead optimization. *Proteins Struct. Funct. Genet.* 30: 321–336.

Patterson D.E., Cramer R.D., Ferguson A.M., Clark R.D. & Weinberger L.E. 1996. Neighborhood behavior: A useful concept for validation of "molecular diversity" descriptors. *J. Med. Chem.* 39: 3049–3059.

Petersen H.G. 1995. Accuracy and efficiency of the particle mesh Ewald method. *J. Chem. Phys.* 103: 3668.

Petukh M., Stefl S. & Alexov E. 2013. The role of protonation states in ligand-receptor recognition and binding. *Curr. Pharm. Des.* 19: 4182–4190.

Phillips J.C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kalé L. & Schulten K. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26: 1781–1802.

Plewczynski D., Łaźniewski M., Augustyniak R. & Ginalski K. 2011. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J. Comput. Chem.* 32: 742–755.

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

Rastelli G., Del Rio A., Degliesposti G. & Sgobba M. 2010. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.* 31: 797–810.

Ravitz O., Zsoldos Z. & Simon A. 2011. Improving molecular docking through eHiTS' tunable scoring function. *J. Comput. Aided. Mol. Des.* 25: 1033–1051.

Ripphausen P., Nisius B. & Bajorath J. 2011. State-of-the-art in ligand-based virtual screening. *Drug Discov. Today* 16: 372–376.

Ripphausen P., Nisius B., Peltason L. & Bajorath J. 2010. Quo vadis, virtual screening? A comprehensive survey of prospective applications. *J. Med. Chem.* 53: 8461–8467.

Roberts B.C. & Mancera R.L. 2008. Ligand-protein docking with water molecules. *J. Chem. Inf. Model.* 48: 397–408.

Rogers D. & Hahn M. 2010. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* 50: 742–754.

Rose P.W., Prlic A., Bi C., Bluhm W.F., Christie C.H., Dutta S., Green R.K., Goodsell D.S., Westbrook J.D., Wools J., Young J., Zardecki C., Berman H.M., Bourne P.E. & Burley S.K. 2015. The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.* 43: 345–356.

Rush T.S., Grant J.A., Mosyak L. & Nicholls A. 2005. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* 48: 1489–1495.

Ryckaert J.P., Ciccotti G. & Berendsen H.J.C. 1977. Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes. *J. Comput. Phys.* 23: 327–341.

Sastry G.M., Adzhigirey M., Day T., Annabhimoju R. & Sherman W. 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided. Mol. Des.* 27: 221–234.

Sastry M., Lowrie J.F., Dixon S.L. & Sherman W. 2010. Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments. *J. Chem. Inf. Model.* 50: 771–784.

Schuster D., Nashev L.G., Kirchmair J., Laggner C., Wolber G., Langer T. & Odermatt A. 2008. Discovery of nonsteroidal 17beta-hydroxysteroid dehydrogenase 1 inhibitors by pharmacophore-based screening of virtual compound libraries. *J. Med. Chem.* 51: 4188–99.

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

Sgobba M., Caporuscio F., Anighoro A., Portioli C. & Rastelli G. 2012. Application of a post-docking procedure based on MM-PBSA and MM-GBSA on single and multiple protein conformations. *Eur. J. Med. Chem.* 58: 431–440.

Sheinerman F.B. & Honig B. 2002. On the role of electrostatic interactions in the design of protein-protein interfaces. *J. Mol. Biol.* 318: 161–177.

Sheinerman F.B., Norel R. & Honig B. 2000. Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.* 10: 153–159.

Shelley J.C., Cholleti A., Frye L.L., Greenwood J.R., Timlin M.R. & Uchimaya M. 2007. Epik: A software program for pKa prediction and protonation state generation for drug-like molecules. *J. Comput. Aided. Mol. Des.* 21: 681–691.

Sheridan R.P. 2008. Alternative global goodness metrics and sensitivity analysis: Heuristics to check the robustness of conclusions from studies comparing virtual screening methods. *J. Chem. Inf. Model.* 48: 426–433.

Sheridan R.P. & Kearsley S.K. 2002. Why do we need so many chemical similarity search methods? *Drug Discov. Today* 7: 903–911.

Shiau A.K., Barstad D., Loria P.M., Cheng L., Kushner P.J., Agard D.A. & Greene G.L. 1998. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* 95: 927–937.

Simões C.J. V, Mukherjee T., Brito R.M.M. & Jackson R.M. 2010. Toward the discovery of functional transthyretin amyloid inhibitors: Application of virtual screening methods. *J. Chem. Inf. Model.* 50: 1806–1820.

Sulea T., Cui Q. & Purisima E.O. 2011. Solvated interaction energy (SIE) for scoring protein-ligand binding affinities. 2. benchmark in the CSAR-2010 scoring exercise. *J. Chem. Inf. Model.* 51: 2066–2081.

Sulea T., Hogues H. & Purisima E.O. 2012. Exhaustive search and solvated interaction energy (SIE) for virtual screening and affinity prediction. In *Journal of Computer-Aided Molecular Design*. pp. 617–633.

Svensson F., Karlén A. & Sköld C. 2012. Virtual screening data fusion using both structure- and ligand-based methods. *J. Chem. Inf. Model.* 52: 225–32.

Tamura H., Tameishi K., Yamada A., Tomita M., Matsuo Y., Nishikawa K. & Ikezawa H. 1995. Mutation in aspartic acid residues modifies catalytic and haemolytic activities of Bacillus cereus sphingomyelinase. *Biochem. J.* 309 (Pt 3: 757–764.

Tan L., Geppert H., Sisay M.T., Gütschow M. & Bajorath J. 2008. Integrating structure- and ligand-based virtual screening: Comparison of individual, parallel, and fused molecular docking and similarity search calculations on multiple targets. *ChemMedChem* 3: 1566–1571.

Tao L., Zhang P., Qin C., Chen S.Y., Zhang P., Chen Z., Zhu F., Yanga S.Y., Wei Y.Q. & Chen Y.Z. 2015. Recent progresses in the exploration of machine learning methods as in-silico ADME prediction tools. *Adv. Drug Deliv. Rev.* 86: 83–100.

Therrien E., Weill N., Tomberg A., Corbeil R.C., Lee D. & Moitessier N. 2014. Docking ligands into flexible and solvated macromolecules. 7. Impact of protein flexibility and water molecules on docking-based virtual screening accuracy. *J. Chem. Inf. Model.* 54: 3198–3210.

Thilagavathi R. & Mancera R.L. 2010. Ligand-protein cross-docking with water molecules. *J. Chem. Inf. Model.* 50: 415–421.

Thompson D.C., Humblet C. & Joseph-McCarthy D. 2008. Investigation of MM-PBSA rescoring of docking poses. *J. Chem. Inf. Model.* 48: 1081–1091.

Totrov M. & Abagyan R. 2008. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* 18: 178–184.

Truchon J.F. & Bayly C.I. 2007. Evaluating virtual screening methods: Good and bad metrics for the "early recognition" problem. *J. Chem. Inf. Model.* 47: 488–508.

Tsui V. & Case D.A. 2000. Molecular dynamics simulations of nucleic acids with a generalized Born solvation model. *J. Am. Chem. Soc.* 122: 2489–2498.

Urbaczek S., Kolodzik A. & Rarey M. 2014. The valence state combination model: a generic framework for handling tautomers and protonation states. *J. Chem. Inf. Model.* 54: 756–766.

Vainio M.J., Puranen J.S. & Johnson M.S. 2009. ShaEP: molecular overlay based on shape and electrostatic potential. *J. Chem. Inf. Model.* 49: 492–502.

Wallach I. & Lilien R. 2011. Virtual decoy sets for molecular docking benchmarks. *J. Chem. Inf. Model.* 51: 196–202.

Wang J., Wang W., Kollman P.A. & Case D.A. 2005. Antechamber, an accessory software package for molecular mechanical calculations. *J. Comput. Chem.* 25: 1157–1174.

Wang J., Wolf R.M., Caldwell J.W., Kollman P.A. & Case D.A. 2004. Development and testing of a general Amber force field. *J. Comput. Chem.* 25: 1157–1174.

Wang Y.T., Su Z.Y., Hsieh C.H. & Chen C.L. 2009. Predictions of binding for dopamine D2 receptor antagonists by the SIE method. *J. Chem. Inf. Model.* 49: 2369–2375.

Warren G.L., Andrews C. V, Capelli A., Clarke B., LaLonde J., Lambert M.H., Lindvall M., Nevins N., Semus S.F., Senger S., Tedesco G., Wall I.D., Woolven J.M., Peishoff C.E. & Head M.S. 2006. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* 5912–5931.

van de Waterbeemd H. & Gifford E. 2003. ADMET in silico modelling: towards prediction paradise? *Nat. Rev. Drug Discov.* 2: 192–204.

Venkatachalam C.M., Jiang X., Oldfield T. & Waldman M. 2003. LigandFit: A novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graph. Model.* 21: 289–307.

Venkatraman V., Pérez-Nueno V.I., Mavridis L., Ritchie D.W., Perez-Nueno V.I., Mavridis L. & Ritchie D.W. 2010. Comprehensive comparison of ligand-based virtual screening tools against the DUD data set reveals limitations of current 3D methods. *J. Chem. Inf. Model.* 50: 2079–2093.

Verdonk M.L., Cole J.C., Hartshorn M.J., Murray C.W. & Taylor R.D. 2003. Improved protein-ligand docking using GOLD. *Proteins Struct. Funct. Genet.* 52: 609–623.

Verma J., Khedkar V.M. & Coutinho E.C. 2010. 3D-QSAR in drug design--a review. *Curr. Top. Med. Chem.* 10: 95–115.

Virtanen S.I. & Pentikäinen O.T. 2010. Efficient virtual screening using multiple protein conformations described as negative images of the ligand-binding site. *J. Chem. Inf. Model.* 50: 1005–1011.

Wolber G., Dornhofer A. a. & Langer T. 2006. Efficient overlay of small organic molecules using 3D pharmacophores. *J. Comput. Aided. Mol. Des.* 20: 773–788.

Wolber G. & Langer T. 2005. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* 45: 160–169.

Vuorinen A., Engeli R., Meyer A., Bachmann F., Griesser U.J., Schuster D. & Odermatt A. 2015. Ligand-based pharmacophore modeling and virtual screening for the discovery of novel 17β-hydroxysteroid dehydrogenase 2 inhibitors. *J. Med. Chem.* 57: 5995–6007.

Yang S.Y. 2010. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov Today* 15: 444–450.

Ylilauri M. & Pentikäinen O.T. 2013. MMGBSA as a tool to understand the binding affinities of filamin-peptide interactions. *J. Chem. Inf. Model.* 53: 2626–2633.

Zhang Q. & Muegge I. 2006. Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: Ranking, voting, and consensus scoring. *J. Med. Chem.* 49: 1536–1548.

Zhu T., Cao S., Su P.-C., Patel R., Shah D., Chokshi H.B., Szukala R., Johnson M.E. & Hevener K.E. 2013. Hit identification and optimization in virtual screening: practical recommendations based on a critical literature analysis. *J. Med. Chem.* 56: 6560–72.

Zsoldos Z., Reid D., Simon A., Sadjad S.B. & Johnson A.P. 2007. eHiTS: a new fast, exhaustive flexible ligand docking system. *J. Mol. Graph. Model.* 26: 198–212.

ORIGINAL PAPERS

I

COMPARISON OF VIRTUAL HIGH-THROUGHPUT SCREENING METHODS FOR THE IDENTIFICATION OF PHOSPHODIESTERASE-5 INHIBITORS

by

Sanna Niinivehmas, Salla Virtanen, Jukka Lehtonen, Pekka Postila &
Olli T. Pentikäinen, 2011

Journal of Chemical Information and Modeling 51: 1353-1363

II

IDENTIFICATION OF ESTROGEN RECEPTOR ALPHA LIGANDS WITH VIRTUAL SCREENING TECHNIQUES

by

Sanna Niinivehmas, Elangovan Manivannan, Sanna Rauhamäki,
Juhani Huuskonen & Olli T. Pentikäinen

Submitted manuscript

III

ULTRAFAST PROTEIN STRUCTURE-BASED VIRTUAL SCREENING WITH PANTHER

by

Sanna Niinivehmas, Kari Salokas, Sakari Lätti, Hannu Raunio & Olli T. Pentikäinen, 2015

Computer-Aided Molecular Design 10: 989-1006

IV

CASE-SPECIFIC PERFORMANCE OF MMPBSA, MMGBSA, AND SIE IN VIRTUAL SCREENING

by

Salla Virtanen, Sanna Niinivehmas & Olli T. Pentikäinen, 2015

In press doi:10.1016/j.jmgm.2015.10.012

V

RELIABILITY OF VIRTUAL SCREENING METHODS IN PREDICTION OF PDE4B-INHIBITOR ACTIVITY

by

Victoria Shubina, Sanna Niinivehmas & Olli T. Pentikäinen, 2015

Current Drug Discovery Technologies 12: 117-26

VI

STRUCTURE-ACTIVITY RELATIONSHIP OF SPHINGOMYELIN ANALOGS WITH SPHINGOMYELINASE FROM BACILLUS CEREUS

by

Christian Sergelius, Sanna Niinivehmas, Terhi Maula, Mayuko Kurita, Shou Yamaguchi, Tetsuya Yamamoto, Shigeo Katsumura, Olli T. Pentikäinen & J. Peter Slotte, 2012

Biochimica et Biophysica Acta 1818: 474-480