JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

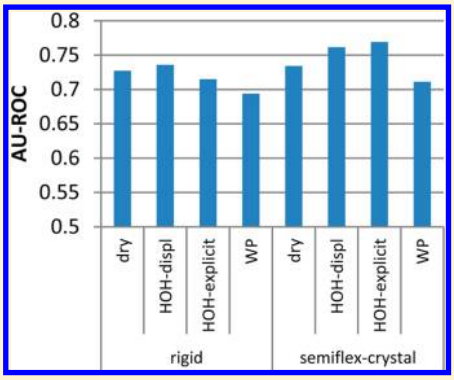
# Docking Ligands into Flexible and Solvated Macromolecules. 7. Impact of Protein Flexibility and Water Molecules on Docking-Based Virtual Screening Accuracy

Eric Therrien, Nathanael Weill, Anna Tomberg, Christopher R. Corbeil, Devin Lee, and Nicolas Moitessier*

Department of Chemistry, McGill University, 801 Sherbrooke Street W., Montréal, Québec, Canada H3A 0B8

Supporting Information

**ABSTRACT:** The use of predictive computational methods in the drug discovery process is in a state of continual growth. Over the last two decades, an increasingly large number of docking tools have been developed to identify hits or optimize lead molecules through in-silico screening of chemical libraries to proteins. In recent years, the focus has been on implementing protein flexibility and water molecules. Our efforts led to the development of FITTED first reported in 2007 and further developed since then. In this study, we wished to evaluate the impact of protein flexibility and occurrence of water molecules on the accuracy of the FITTED docking program to discriminate active compounds from inactive compounds in virtual screening (VS) campaigns. For this purpose, a total of 171 proteins cocrystallized with small molecules representing 40 unique enzymes and receptors as well as sets of known ligands and decoys were selected from the Protein Data Bank (PDB) and the Directory of Useful Decoys (DUD), respectively. This study revealed that implementing displaceable crystallographic or computationally placed particle water molecules and protein flexibility can improve the enrichment in active compounds. In addition, an informed decision based on library diversity or research objectives (hit discovery vs lead optimization) on which implementation to use may lead to significant improvements.




## INTRODUCTION

Traditional medicinal chemistry has been instrumental to developing a number of drugs. However, for the past decades, constant advances in robotics, spectroscopy, computer science, and hardware have introduced novel strategies for drug design. Nowadays, high throughput screening (HTS), fragment-based drug design (FBDD),[1] and computational drug design techniques are among the most popular approaches. More recently, HTS has been increasingly complemented by virtual screening (VS), its in-silico counterpart.[2−6] In fact, the combination of improved predictive computational approaches, development of large databases of small molecules, and ever increasing CPU (and GPU) power has led us to a point where virtual chemical libraries can be assembled in a few clicks and virtually tested for bioactivity.[7] In addition, the number of available protein and nucleic acid structures that can be used for 3D structure-based VS is currently growing at a tremendous rate with nearly 100 000 protein structures housed at the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) which itself is growing with thousands of proteins being added each year.[8] As a result, VS of medium to large libraries of compounds has become a routine practice to quickly increase hit rates in lead discovery. When undertaking a lead discovery program, choosing between two methods (or a combination thereof) such as HTS or VS is of great importance as the associated costs, required time, and personnel are significantly different. Although the accuracy of predictions is never perfect, virtually screening the compounds offers a number of benefits. A VS study is easier to setup, faster to run, cheaper, and does not suffer from low purity of samples, variability of the protein samples, and other experimental uncertainty. However, VS techniques have also their own limitations resulting from a number of approximations (i.e., protein models in vacuo and in the absence of salts or other macromolecules).

Structure-based VS generally involves docking of a library of small molecules to a protein structure followed by a ranking of these molecules by predicted binding affinities.[9] In practice, properly docking a ligand into a protein depends on a wide variety of factors. Efficient conformational search of the possible ligand binding modes, and a proper identification and use of interactions of the ligand with the protein residues (and metals where applicable) within the binding site are two examples.[10] As soon as a pose is generated, it needs to be evaluated. Only the best one is retained for each compound. Then all these "scored" compounds should be compared. For this purpose, scoring functions have been developed to predict binding affinities and/or to place the most likely active compounds at the top of the list.[11] At this stage, both

Received: May 16, 2014
Published: October 3, 2014

Table 1. Selected Protein/Receptor Structures and Ligand Sets[a]

| protein/receptor | acronym | PDB codes | no. actives | diversity[b] | no. decoys |
|---|---|---|---|---|---|
| Nuclear Hormone Receptors | | | | | |
| androgen receptor | AR | **2ao6**[c], 2ax9, 2axa, 2pnu, 3b68 | 40 | 0.606 | 960 |
| estrogen receptor agonist | ERagonist | 1gwq, 1gwr, **1l2i**, 1xpc, 2iog | 40 | 0.548 | 960 |
| estrogen receptor antagonist | ERantagonist | 1xp1, 2iok, 2ouz, 3dt3, **3ert** | 39 | 0.681 | 936 |
| glucocorticoid receptor | GR | **1m2z** | 40 | 0.709 | 960 |
| mineralocorticoid receptor | MR | 2a3i, **2aa2**, 2aa5, 2aa7 | 15 | 0.743 | 360 |
| peroxisome proliferator-activated receptor gamma | PPARg | **1fm9**, 2hwq, 2q61, 2q8s, 2zno | 40 | 0.789 | 960 |
| progesterone receptor | PR | **1sr7**, 1zuc, 2w8y, 3d90, 3g8n | 27 | 0.693 | 648 |
| retinoid X receptor alpha | RXRa | 1mv9, **1mvc**, 2p1v, 3fc6, 3fug | 20 | 0.859 | 480 |
| Kinases | | | | | |
| cyclin-dependent kinase 2 | CDK2 | **1ckp**, 1h0v, 2r3o, 2vti, 2vto | 40 | 0.541 | 960 |
| epidermal growth factor receptor | EGFr | **1m17**, 1xkk, 2gs6, 2rgp, 3bel | 40 | 0.633 | 960 |
| fibroblast growth factor receptor 1 | FGFr1 | **1agw**, 1fgi, 2fgi | 40 | 0.646 | 960 |
| heat shock protein 90 | HSP90 | **1uy6**, 1uyd, 2bz5, 2uwd, 3ekr | 24 | 0.669 | 576 |
| p38 mitogen-activated protein | P38-MAP | **1kv2**, 1oz1, 1wbo, 1wbv, 2rg6 | 40 | 0.611 | 960 |
| platelet derived growth factor receptor beta | PDGFrb | homology model | 40 | 0.562 | 960 |
| proto-oncogene tyrosine-protein kinase Src | SRC | 1y57, 1yol, 2bdj, **2src**, 2oiq | 40 | 0.631 | 960 |
| thymidine kinases | TK | 1e2n, 1ki2, 1ki7, **1kim**, 1e2k | 22 | 0.808 | 528 |
| vascular endothelial growth factor receptor 2 | VEGFr2 | 1ywn, **2xir**[d] | 40 | 0.553 | 960 |
| Serine Proteases | | | | | |
| factor Xa | FXa | **1f0r**, 1f0s, 2p16, 2uwl, 2xbw | 40 | 0.756 | 960 |
| thrombin | THROMBIN | **1ba8**, 1gj5, 1lhg, 2zff, 2zgb | 40 | 0.701 | 960 |
| trypsin | TRYPSIN | **1bju**, 1gi6, 1tnk, 1utp, 1y3u | 40 | 0.759 | 960 |
| Metalloenzymes | | | | | |
| angiotensin-converting enzyme | ACE | 1o8a, **1o86**, 1uze, 1uzf, 2oc2 | 40 | 0.655 | 960 |
| adenosine deaminase | ADA | **1ndw**, 1o5r, 1v7a, 1wxy, 2z7g | 23 | 0.645 | 552 |
| catechol-o-methyltransferase | COMT | **1h1d**, 1vid, 2cl5, 2zvj | 11 | 0.487 | 288 |
| phosphodiesterase 5A | PDE5 | 1t9s, 1tbf, 1xoz, **1xp0** | 40 | 0.591 | 960 |
| Folate Enzymes | | | | | |
| dihydrofolate reductase | DHFR | **3dfr** | 40 | 0.704 | 960 |
| GAR transformylase | GART | **1c2t**, 1c3e, 1cde, 1gar, 1jkx | 21 | 0.749 | 504 |
| Other Enzymes | | | | | |
| acetylcholinesterase | AChE | **1eve**, 1gpn, 1zgc, 2ack, 2ckm | 40 | 0.720 | 960 |
| aldose reductase | ALR2 | 1ah0, **1ah3**, 1eko | 26 | 0.545 | 624 |
| AmpC $\beta$-lactamases | AmpC | 1iel, 1iem, 1l2s, 1llb, **1xgj** | 21 | 0.704 | 504 |
| cyclooxygenase 1 | COX-1 | 1eqg, 1igz, **1q4g**, 2ayl, 2oye | 25 | 0.605 | 600 |
| cyclooxygenase 2 | COX-2 | **1cx2**, 3pgh, 4cox, 6cox | 40 | 0.689 | 960 |
| glycogen phosphorylase B | GPB | **1a8i**, 1ggn, 1p4g, 2qn7, 2qn8 | 40 | 0.666 | 960 |
| HIV-1 protease | HIVPR | 1dif, 1hih, **1hpx**, 1w5w, 2qnp | 40 | 0.696 | 960 |
| HIV-1 reverse transcriptase | HIVRT | 1rth, **1rt1**, 1vrt, 1vru, 1c1b | 40 | 0.540 | 960 |
| HMG-CoA reductase | HMGR | **1hw8**, 1hwi, 3ccw, 3ccz, 3cdb | 35 | 0.747 | 840 |
| enoyl reductase | InhA | **1p44**, 2b35, 2h7l, 2h7n, 2x23 | 40 | 0.649 | 960 |
| neuraminidase | NA | **1a4g**, 1a4q, 1nsd | 40 | 0.691 | 960 |
| poly(ADP-ribose) polymerase | PARP | **1efy**, 1pax, 2pax, 3pax, 4pax | 33 | 0.821 | 792 |
| purine nucleoside phosphorylase | PNP | 1b8n, **1b8o**, 2ai1, 2ai2, 3fuc | 25 | 0.656 | 600 |
| S-adenosyl-homocysteine hydrolase | SAHH | **1a7a** | 33 | 0.831 | 792 |

[a]The pdb codes of the crystal structures provided by the DUD are given in bold. [b]Average Tanimoto coefficient between all active compounds computed using SELECT. [c]2ao6 replaced 1xq2 initially reported.[36] [d]1vr2 is an apo structure and was replaced by 1ywn.

knowledge of the protein targets and knowledge of the docking tools used to study them are of paramount importance. Some proteins are more flexible than others, and in such cases, it stands to reason that a program that considers protein flexibility reliably should be used over a program that does not. As a result, docking programs have been modified to search for optimal protein conformations.[12−14] Handling of conserved waters is another factor that needs to be addressed,[15,16] and implementations were proposed.[17−19] As a result, some programs now include both water molecules and protein flexibility[10] including AutoDock 4.2,[19,20] FlexX (with FlexX-Ensemble),[21] GOLD,[22] and PLANTS.[11,23,24] However, one question arises: are these additional implementations (protein flexibility and water molecules) increasing the accuracy of the programs or simply adding noise to the predictions? Some reports with either a single program,[25] a few protein classes, and even the complete DUD set[26] indicated that protein flexibility improves the accuracy.[27,28] The present work attempts to provide a more general answer to this key question and to complement our previous two studies on the effect of water molecules and protein flexibility on docking program accuracy. Our previous studies evaluated the impact of these factors on

pose prediction[29] and scoring function accuracy[30] while the current study will look at their impact on docking-based VS campaign success rate. In addition, this manuscript reports our novel implementations for generating protein structures and locating water molecules. While this manuscript was in preparation, a report from Schumann and Armen describing a novel method for docking to flexible proteins revealed that considering protein flexibility can improve the VS outcome in a number of cases.[31]

**Background.** We have been developing FITTED, a docking tool with a major focus on protein flexibility and water molecules.[29,30,32−35] As expected, implementing protein flexibility and water molecules were found to slightly improve the accuracy of the binding modes proposed by FITTED. Our previous validation study of docking programs for their ability to predict the binding modes of ligands has shown that most of the major docking programs were affected by protein flexibility and to less extent (if at all) by the presence of critical (e.g., conserved) water molecules.[29]

While a study such as this one can show the strengths and weaknesses of our program, ultimately the effectiveness of prospective VS will depend on many other factors (e.g., availability of libraries and target structure, resolution of the target structure, number of true active compounds in the libraries) unique to this particular screening. In the present study, our docking program FITTED was applied to specifically selected proteins and small molecules. The key focal points of this study included: the effects of protein flexibility using either a set of crystallographic or computationally derived structures, the effects of conserved water molecules in the binding site, the inclusion of water particles, and the transferability to various proteins. The results and conclusions from this study will provide valuable information to end users of docking tools as well as valuable data for the developers of these tools for future improvements and refinements. It is hoped that this study will also further the general knowledge in docking-based VS and guide future and current users of VS tools to decide what conditions are most appropriate to use for any given target and/or application.

## METHODS

**Ligand Data Sets.** The accuracy of a VS tool is traditionally evaluated using a set of known bioactive compounds (referred to as actives) combined to a set of bioinactive compounds (referred to as inactives or decoys). The decoys should resemble the active ligands in physicochemical properties so that the enrichment is not simply a separation of obvious physical features (i.e., molecular size, charge), but at the same time they must also be chemically distinct from the actives to ensure that they are likely inactive. The Directory of Useful Decoys (DUD)[36] was made to address this issue and reduce this bias. Despite the known issues of this set such as decoys found to be active later on,[37] it is expected to provide useful information in the context of this study. Herein we are comparing different settings (relative accuracy) with the same sets rather than looking at absolute accuracy. An enhanced version of this database came out (DUD-E)[38] while this study was ongoing, but was not used in this work. Thus, the DUD set is suitably challenging for evaluating the ability to rank order active compounds over inactive compounds (actives over decoys). In addition, this set has been widely used for program evaluation and investigation purposes.[26,39]

In this work, the DUD ligands and decoys sets from each of the 40 targets were retrieved and prepared as follows. The DUD actives and decoys libraries were first curated to remove any duplicates and errors in the molecular structures. For example, several compounds from the Glucocorticoid Receptor (GR, Table 1) set where missing bonds and incomplete structures were originally provided. In addition, amidine groups present in thrombin and trypsin ligands were incorrectly protonated due to incorrect bond order in their structures. We then used a random selection to obtain subsets derived from the cleaned DUD sets with molecules unique in structure and more than one tautomer present if needed. Due to CPU and time constraints (millions of docking runs have been necessary to complete this study), we decided to include a maximum of 1000 distinct molecules in total for each target, keeping a constant ratio of 1:24 (e.g., 40 actives/960 decoys). In cases where fewer than 40 active compounds were available, the number of decoys selected was adjusted accordingly (Table 1). In order to ensure that the selection of compounds was not detrimental to the diversity of the data set, clustering was carried out using our program SELECT described previously.[7] In a nutshell, SELECT describes each molecule using MACSS keys, measures similarity using the Tanimoto coefficient and clusters compounds by similarity using a K-means approach.[7] The recursive clustering is carried out until the above-mentioned number of compounds was reached.

The small molecules were then prepared for docking using SMART, a previously reported program, member of the FORECASTER Suite.[7,29,32]

**Selection of Protein Structures.** The DUD includes a single crystal structure for 39 targets (pdb code in bold in Table 1) and a homology model for PDGFrb. Including more than one structure for each enzyme was a key element for the determination of the effect of different water molecules, receptor flexibility and protein structure. Additional structures were therefore required. In order to carefully choose these additional PDB structures, we used a cluster sequence similarity cutoff of 100% of the reference PDB structure (Table 1). When the number of available protein structures exceeded five, we applied specific criteria to select five representative structures. First, low resolution structures were removed (resolution >2.5 Å), then the five most dissimilar cocrystallized ligands were identified using our program SELECT. Finally, when some of the five selected structures were not satisfying (i.e., no cocrystallized waters), manual selection was carried out.[30] Finally, the number of unique crystal structures varied from one to five structures per protein (Table 1).

**Docking Program.** FITTED docks the ligands into flexible proteins using a matching algorithm-enhanced genetic algorithm.[29,32] Previous reports from our group detailed the development of previous FITTED versions and only a brief description is given below.[29,32,33] Protein side-chains, protein backbone conformations, and ligand torsion angles as well as water molecule (or particle) positions are all optimized through a Lamarckian/Darwinian evolutionary process.[32] Each protein/water (or water particle)/ligand complex is described using a single chromosome on which evolution operators are applied. To select poses, a complex consensus docking based on various scores of increasing precisions described previously is used. Not only can the waters (or particles) move through the evolution but they can also be displaced during the docking/scoring process. Displacement is considered through the use of a specifically designed energy function which turns off water

Table 2. Acronyms and Abbreviations Used Throughout the Manuscript

| mode | acronym | water or protein flexibility | description |
|---|---|---|---|
| protein flexibility mode | rigid | rigid protein docking | a single conformation of the protein is used while docking small molecules |
| | flex | fully flexible protein docking mode | input protein conformations will be used to generate composite structures |
| | semiflex | semiflexible protein docking mode | small molecules will be docked to static conformational ensembles |
| protein structures | crystal | experimental protein conformations | crystal structures are used by the docking program |
| | side-chain | newly generated protein conformations | a side-chain rotamer library is used to generate new protein conformations |
| | tc | newly generated protein conformations | whole protein flexibility is considered using tCONCOORD |
| water mode | dry | no water molecules | water molecules removed prior to docking |
| | HOH-crystal | crystallographic water molecules | these water molecules are kept throughout the docking runs (explicit nondisplaceable water molecules) |
| | HOH-displ | crystallographic water molecules | these explicit all-atom water molecules can be displaced throughout the docking runs |
| | WP | spherical water particles | these single "atom" water molecules can be displaced throughout the docking runs |
| | HOH-to-WP | crystallographic water molecules assigned WP parameters | these single atom water molecules can be displaced throughout the docking runs |

molecules overlaying with any ligand atom. The docking poses are then evaluated with the implemented RankScore scoring function.[34] RankScore is a previously reported scoring function[34] combining force field terms (van der Waals, electrostatic and hydrogen bonding interactions), solvation terms (from GB/SA method[40]) and a ligand entropy term. The latter penalizes for rotatable bonds that would be frozen upon binding, considers polarity of the direct chemical environment of each bond as well as their burying in the protein.

**Considering Protein Flexibility through Conformational Ensembles.** Our docking program FITTED considers protein flexibility using a conformational ensemble of protein structures. These conformational ensembles can either be a set of experimental structures or a set of conformations produced through computation. The flexibility of proteins can be somewhat uncoupled into two kinds of motions; either the side-chains can adopt different conformations and/or the backbone can fold slightly differently. To investigate these two factors, three different approaches were tested. All the protein structures preparation described herein has been implemented into and performed using our automated programs MATCH-UP and PREPARE.[7]

The first approach consisted in using the above-listed crystal structures as conformational ensembles. With these experimental structures, the backbone and the side-chains were considered flexible when docking small molecules.

The second approach used libraries of side-chains rotamers. A recent implementation in the PREPARE program produces protein conformations by evaluation of side-chains conformations. According to the statistical analysis carried out by Najmanovich and co-workers,[41] some residues are more flexible than others. Thus, the five most likely flexible residues in the binding site were identified and a maximum of three of these five residues were moved simultaneously for a possibility of ten possible residue triades. The conformation of each of these side chains was systematically searched. For this purpose, Lovell's library of rotamers was used.[42] As the energy well of side-chain conformations can hardly be represented by a small set of conformations,[43] we next refined these side-chains. For example, an arginine side-chain is assigned all of the 34 possible rotamers listed in this library. Each of these 34 conformations is further optimized by systematically rotating each dihedral angle by either 0, 2.5, or 5.0° clockwise or counterclockwise and selecting the most favored conformation. The selection of this preferred conformation is based on the potential energy as computed using the molecular mechanics routines previously implemented in PREPARE. When this search is complete, the four most energetically favored protein conformations (as computed by PREPARE) together with the native structure are archived. With this implementation, the backbone has a single conformation in all structures and will consequently be considered rigid while the side-chains will be considered flexible.

The third approach to mimic protein flexibility was based on an ensemble of conformations generated from a single crystal structure. This can be achieved using tCONCOORD program, which is a geometry-based approach to sample proteins conformational space.[44,45] In contrast to simulation techniques such as molecular dynamics, the protocol implemented in tCONCOORD is very fast and computationally undemanding. First, all pairwise interatomic distances are measured in the input structure and a set of geometrical constraints, describing the upper and lower bounds of these distances, is produced. This geometrical map of the protein is used to generate new conformations of the original structure. Starting from random coordinates, atoms are iteratively repositioned in order to minimize constraints' violations. This procedure is repeated until the sum of violations is zero. This approach reduces the conformational space of atomic coordinates but does not guarantee the produced conformation is realistic in terms of energy. For this reason, HOPPscore software[46] was used to identify five structures with the highest score. HOPPscore provides a convenient way of assessing the quality of protein structures based on higher-order $\phi-\psi$ maps. Similarly to how a Ramachandran plot discriminates the torsions in a single residue to favorable and unfavorable regions, higher-order $\phi-\psi$ maps describe the allowed and disallowed torsions for short fragments of the structure leading to a multidimensional plot.[47] Using these two tools we selected the five structures with best energy computed for flexible docking in FITTED (Table 2).

All these protein structures (crystal structures, structures reconstructed using rotamer libraries and structures generated from tCONCOORD) were further processed: hydrogen atoms were added when needed, rotamers (e.g., serine hydroxyl groups, asparagine terminal amides), tautomers, protonation states (e.g., histidine residues), and orientation of water molecules were optimized, and missing side chains were reconstructed as described previously.[7] This optimization relied on molecular mechanics routines computing the energy of each conformation using the Generalized Amber Force Field.[48] In

order to simulate protein flexibility[29,32,34] with these protein conformational ensembles, these multiple protein structures were automatically superposed using MATCH-UP before being processed.

The protein structures obtained through these different protocols were next prepared for docking using PROCESS. In particular, PROCESS labels any residue with at least one atom within 7 Å from the cocrystallized ligands as being flexible. This labeling will later be used by FITTED to identify which residues should be considered flexible.

**Water Molecules and Water Particles.** Two approaches were used herein to evaluate the impact of water molecules on docking pose prediction and docking-based VS accuracy. In one case, crystallographic water molecules were selected based on their proximity to both protein and ligand atoms, hydrogens were added, and their orientation optimized. In the other case, a method positioning water particles (i.e., united-atom water molecules) was implemented within PREPARE. In contrast to explicit water molecules used in the first approach, water particles are spherical particles that are both hydrogen bond donors and acceptors in an approach similar to that of Forli and Olson.[19] To position these particles, PREPARE relies on an iterative process in which particles are first added at each point of a grid encompassing the binding site with a spacing of 0.25 Å. The particles are then all allowed to move to optimize their interaction with the proteins (the ligands were ignored). The final set of particles is then analyzed and the ten most energetically favored and nonoverlapping (distance between two particle centers greater than 2.8 Å) ones are kept. In preliminary testing on a few test cases, we found that using too many particles was detrimental to the required time for docking. Although the optimized particles were often close to the experimentally observed water molecules, their ranking was not accurate enough to rank the best ones at the top of the list. However, they often appeared in the top ten of the list, hence, selecting ten for the following docking was believed to be optimal. The force field parameters of these neutral particles (either H-bond donor or acceptor) were developed to reproduce both the energy profile of the water molecule hydrogen atoms interacting with either oxygen, nitrogen or sulfur atoms with partial charges of −0.4 and between the water oxygen atoms and other hydrogen bond donors as in explicit water molecules (Figure 1). Although the position of these water particles was refined by molecular mechanics, the entropy is not considered by the current optimization routine thus the accuracy of their placement had to be evaluated. For this purpose, we compared the positions of the water molecules in contact with the cocrystallized ligands observed in crystal structures and the positions of the particles predicted by PREPARE. A particle was found within 1.0 or 2.0 Å of 20% or 41% of the crystalline water molecules. Considering that the placement is critical for optimal pose prediction,[49] we expected these particles to significantly affect the docking accuracy. On one side, these particles may improve the pose when properly placed but may either perturb the binding or may be displaced when not properly positioned. The accuracy we measured for our protocol can hardly be compared with accuracy of other programs using either Monte Carlo (MC) or Molecular Dynamics (MD) simulations such as WaterMap.[50] When condensed phase is used (e.g., MD), it is very likely that a water molecule is found within 2.0 Å of a water molecule observed in crystal structures while our method keeps only a few.

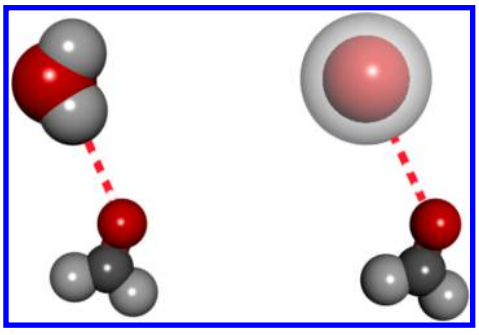


**Figure 1.** Explicit water molecule and particle water.

In our FITTED docking program, the solvation free energy is considered by a Generalized Born/Solvent Accessible Surface Area (GB/SA) implementation instead. Particle waters are treated exactly as are water molecules by the docking program.

**Evaluating the Impact of Protein Flexibility and Water Molecules.** Several different run modes were used in this work for each of the proteins being studied and acronyms will be used throughout this manuscript (Table 2). As described above, three approaches were used to generate conformational ensembles (crystal structures—*crystal*, flexible side-chains—*side-chain*, and remodeled structures using tCONCOORD—*tc*). Within FITTED, the protein is represented as chromosomes with backbone and side-chains as individual genes. If genetic operators such as crossover and mutations are applied to these genes, composite structures will be generated and the protein conformational space will not be restricted to the input conformations. This mode will be referred to as fully flexible protein (*flex*). In contrast, if the input conformations (i.e., conformational ensembles) remain unchanged, the protein flexibility mode is referred to as semiflexible (*semiflex*). Thus, the flexibility of the protein could be considered by different modes (*flex-crystal*, *semiflex-crystal*, *semiflex-side-chain*, and *semiflex-tc*) or not (*rigid*).

The water molecules are also considered in different ways. First, the proteins were considered with all crystallographic waters removed if any, herein described as "dry" proteins. When the crystallographic waters that were found in the binding site of the proteins were kept, they are referred to as "HOH-crystal". FITTED has the ability to displace some water molecules while keeping others during docking; this approach was referred to as "HOH-displ". In addition to explicit water molecules, inclusion of displaceable particle waters was evaluated and referred to as "WP". In order to uncouple the position from the parameters of the water particles and test them individually, we also converted the crystallographic waters into WPs (position of crystallographic water and parameters of WPs, "HOH-to-WP").

## RESULTS AND DISCUSSION

**Metrics for Evaluation of Performances.** While a docking tool should be able to reproduce known binding modes, the true value when conducting a VS is its ability to find unique ligands with potentially novel binding modes. Thus, the performance of docking tool can be first assessed by its ability to reproduce experimentally observed binding modes.[29] To validate all the implementations and ensure the absence of conflicts between routines, self- and cross-docking experiments were first performed and compared to previously reported data. As reported previously, RMSD has been widely used and will be used herein for comparison purposes. For self-docking

experiments, poses with RSMDs below 2.0 will be considered correct. As the average RMSD of protein binding site alpha carbons was 0.60 Å after superposition, the RMSD criterion for accurate cross-docking experiments was reevaluated to 2.60 Å. In the context of VS, the accuracy can be assessed by a set of ligands containing a number of known actives with a number of decoys to rank order the actives over the inactives. An important metric that is commonly used for the evaluation of the performance of ranking methods in VS is the area under the receiver operating characteristic curve (AU-ROC).[51] This metric possesses many desirable statistical behaviors and it is claimed to be independent of the ratio of actives and decoys although this claim has been questioned.[52,53]

**Validation of the Novel Implementations.** In order to evaluate the effect on these different docking modes on pose prediction, a number of docking experiments were carried out as summarized in Table S1. A set of 242 crystal structures were used in self-docking and of 839 complexes in cross-docking. All of the crystal structures selected in Table 1 as well as structures used previously for validating FITTED[29] were included. The ligand and protein structures were prepared using the accessory programs of the FORECASTER[29] suite (SMART, PREPARE, and PROCESS) using default parameters then docking experiments were carried out. The measured accuracy was found to be similar to accuracies of 62% (self-docking), 40% (cross-docking), and 50% (semiflex without cognate structures) recorded with the previously reported version of FITTED (v. 2.6).[29] Although the set has been expanded rendering the numbers difficult to compare, this preliminary study demonstrated the reliability of the current version and validated the more recent implementations. A more complete evaluation is provided as Supporting Information.

**Evaluating a VS Program's Accuracy.** In the most simple terms, the major goal of a study of VS programs is to assess the efficacy of a given program to rank a large database of compounds or to enrich libraries in actives for a given protein target or set of protein targets. Due to the large number of studies of this type, many shortcomings have already been identified.[11,54−56] In fact, VS tools can be exceedingly complex with numerous parameters that must be correctly adjusted for a given set of compounds docked to a specific protein. On the other hand, the testing set may overlap with the protein/ligand set that the program has been trained on or the protein set may be too narrow to be significantly representative. In order to reduce this bias, our program was used with default parameters only and no attempts were made to improve the accuracy through optimization of for example genetic algorithm parameters. During the course of our work, Bauer et al. reported the public DEKOIS 2.0 library that provides new benchmark sets for a variety of targets designed to avoid several issues observed with the bioactivity data.[57]

**This Study.** The objectives of this retrospective study were to evaluate the impact of protein flexibility and water molecules on the accuracy of virtual screens and to provide guidelines for prospective studies. Collecting statistically relevant data required a large number of data points hence large data sets. We are aware that the conclusions may only be the results of our implementations and data set and have tried to reduce these biases. Also, no specific knowledge on the protein targets was used to reduce the impact of this knowledge over the impact of protein flexibility and water molecules.

**First Analysis.** First, preliminary analysis of these thousands of data points revealed that out of the 40 proteins studied, four (COX-1, GR, PDE5, and PDGFrb which is a homology model) were consistently providing inaccurate results (AU-ROC < 0.55) regardless of the docking modes used. Next, the highest AU-ROC measured with ALR2, HIVRT, P38-MAP, and SRC did not exceed a low 0.68. It is interesting to note that the same proteins (ALR2, COX-1, GR, PDE5, PDGFrb, P38-MAP, and SRC) were also leading to poor accuracy in a separate study by Schumann et al.[31] In this later study, another six (ACE, DHFR, FGFr1, HIVPR, HMGA, PR) were also found problematic. In a study by Liebeschuetz et al. using the GOLD program, ALR2, COX-1, GR, and PDGFrb also provided poor enrichments while ICM provided AU-ROC values with PDGFrb, GR below 0.5 with either refined scores or refined protein coordinates while P38-Map, HIVRT, COX-1, and others provided values below 0.70.[58] This first analysis indicated that some proteins (e.g., GR, PDGFrb, P38-MAP, COX-1) are either problematic or providing AU-ROC below 0.70 independently of the program used, revealing that our conclusions would be primarily guided by the system under investigation rather than the nature (ie., methods and algorithms) of our own implementations.

In a third category, screening with AChE, EGFr, FGFr1, PNP, and VEGFr2 was accurate with AU-ROC greater or equal to 0.70 only when both protein flexibility and displaceable crystallographic water molecules were considered. This is a first indication that protein flexibility and water molecules have a positive impact on screening accuracy. As discussed below, docking to the proteins listed as problematic by others might be significantly more accurate if the protein structure selection was performed differently or if scoring functions were further refined. For instance, we have found that docking the DUD ligand set to FGFr1 could be highly accurate (i.e., AU-ROC greater than 0.80) if the right crystal structure was used while poor (i.e., lower than 0.60) if another one is used. Similarly, Liebeschuetz et al. reported AU-ROCs below 0.50 for GR with three scoring functions but a value of 0.77 when ChemScore was used.[58]

To further demonstrate that the conclusions drawn herein are not the result of our specific implementations, we have investigated flexibility using sets of conformations derived differently and water molecules using our two different implementations.

**Impact of Protein Structure and Flexibility.** A careful analysis of the several screens revealed that the selection of the protein structure is critical. Although this was somewhat expected, it appeared to have a stronger effect than initially thought. For example, the measured AU-ROC varies from a poor 0.59 (barely better than random selection) to a high 0.85 using 3pgh and 1cx2, respectively, two crystal structures of the same protein (COX-2). The same large effect has been observed with other proteins as shown in Figure 2.

Among the proteins that have a large variance in the accuracy are kinases (e.g., FGFr1 in Figure 2). Kinases may adopt two major conformations: the active conformation and the inactive conformation which are often referred to as DFG-in and DFG-out, DFG being a motif on the activation loop. In addition to these two major conformations, transition conformations that are significantly different can also be observed.[59] As shown in Figure 3a, a loop featuring Phe489 adopts two significantly different conformations with Lys514 blocking access to an additional pocket in 1fgi but not in 2fgi while an helix bearing Met742 of EGFr adopts two different orientations modulating the space available to inhibitors (Figure 3b). Thus, docking and

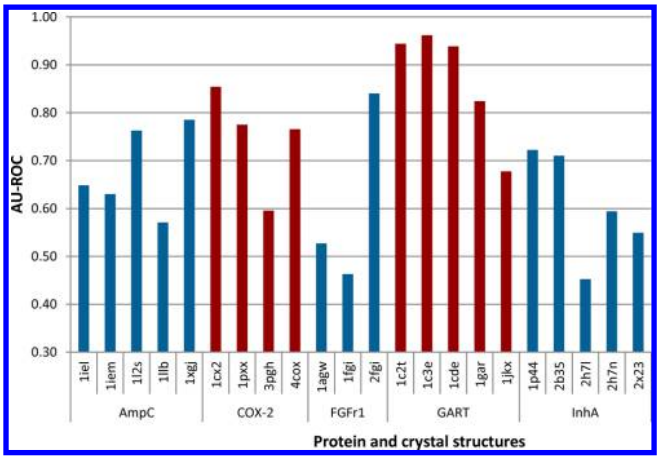



**Figure 2.** Impact of protein structure on VS accuracy (rigid protein).

scoring kinase inhibitors is expected to be very challenging if protein flexibility is not considered.

This large variation of accuracy is highly dependent on the small molecule library docked and may not be transferable to other libraries; 1cx2 might be the "best" structure when docking the DUD set as the selected ligands dock better in this structure but may not accommodate other chemical series appearing in other libraries which may prefer 3pgh.

From this first observation, one can expect that considering protein flexibility while docking should reduce this bias and allow the discovery of novel chemical series in prospective studies. To simulate the two scenarios (i.e., a user uses a single crystal structure or considers protein flexibility), the VS was carried out using the crystal structure suggested on the DUD set and the data was compared to AU-ROC values obtained when the flexibility modes were used. As shown in Figure 4, in about half of the proteins, the protein flexibility mode semiflex and the rigid modes provide similar AU-ROC (within 0.05). However, in 33% of the cases, the protein flexibility (semiflex) provided significantly more accurate results (improvement greater than 0.05) while in 12% was the rigid mode more accurate. With the flex mode, the accuracy was very similar to the rigid protein mode.

The largest increase in AU-ROC when moving from rigid to flexible proteins was observed with FGFr1, HSP90, and PR, and a more in-depth analysis was carried out to understand the cause of this observation (Table 3). Two of the three FGFr1 structures led to AU-ROC below 0.60 including the structure proposed from the DUD. However, the use of 2fgi restored a very good accuracy with AU-ROC above 0.80. In fact, as shown in Figure 3, 2fgi has an open pocket while the other two

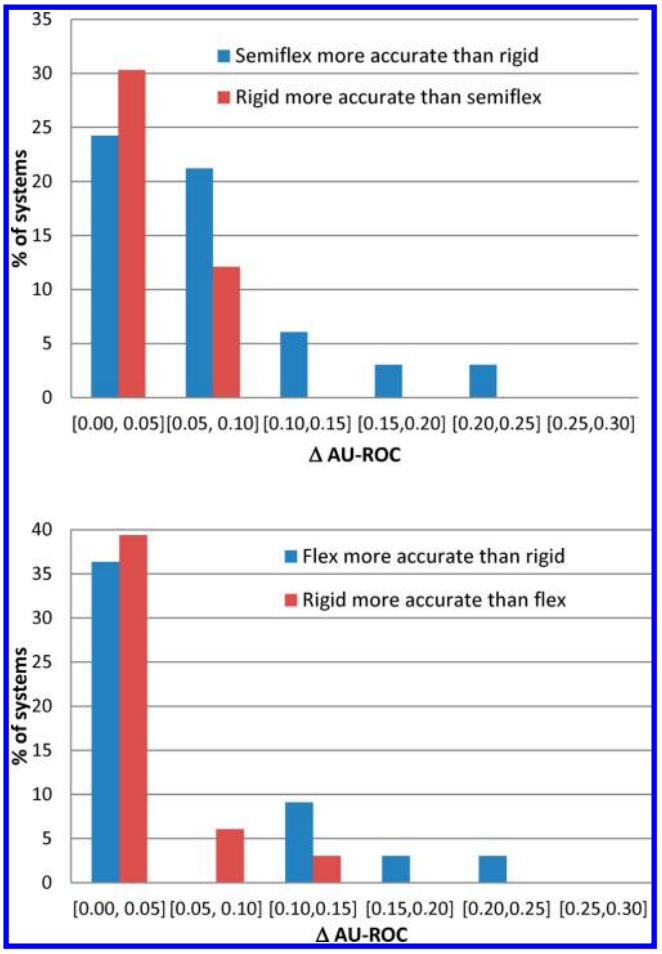



**Figure 4.** VS using either flexible or rigid proteins (HOH-displ): (top) semiflex mode, (bottom) flex mode.

structures have Lys514 blocking the entry to this binding pocket. With the same three structures, the AU-ROC value was 0.66 or 0.70 with the flex and semiflex mode, respectively. In this case, all the active ligands bind well to the 2fgi structure alone and considering protein flexibility adds noise to the calculations.

The largest accuracy loss was observed for PPARγ with AU-ROC decreasing by 0.08 when going from rigid to flexible protein (Table 3). The reason for this drop was very similar to the improved accuracy observed with FGFr1. Out of the five crystal structures in rigid mode, one provided AU-ROC of nearly 0.90 while two provided accuracies as low as 0.62 and 0.65. Since the structure used to compare accuracy of flexible and rigid protein was the most accurate one, adding flexibility is

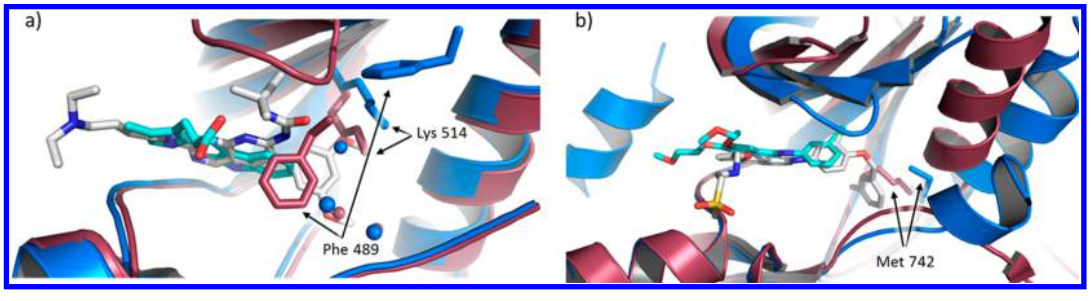



**Figure 3.** (a) 1fgi and 2fgi crystal structures of FGFr1-inhibitor complexes. (b) 1m17 and 1xkk crystal structures of EGFr-inhibitor complexes. The inhibitors and key residues are shown in sticks while key water molecules are shown as spheres.


dx.doi.org/10.1021/ci500299h | J. Chem. Inf. Model. 2014, 54, 3198−3210

Table 3. Accuracy Observed with FGFr1, PR, HSP90 and PPARg (HOH-displ)

| protein | structure used[a] | AUC-ROC |
|---|---|---|
| FGFr1 | **1agw** | 0.55 |
| | 1fgi | 0.44 |
| | 2fgi | 0.81 |
| | flex-crystal | 0.66 |
| | semiflex-crystal | 0.70 |
| PR | **1sr7** | 0.66 |
| | 1zuc | 0.66 |
| | 2w8y | 0.62 |
| | 3d90 | 0.70 |
| | 3g8n | 0.58 |
| | flex-crystal | 0.86 |
| | semiflex-crystal | 0.82 |
| HSP90 | **1uy6** | 0.67 |
| | 1uyd | 0.82 |
| | 2bz5 | 0.78 |
| | 2uwd | 0.69 |
| | 3ekr | 0.74 |
| | flex-crystal | 0.85 |
| | semiflex-crystal | 0.88 |
| PPARg | **1fm9** | 0.89 |
| | 2hwq | 0.68 |
| | 2q61 | 0.63 |
| | 2q8s | 0.65 |
| | 2zno | 0.76 |
| | flex-crystal | 0.74 |
| | semiflex-crystal | 0.80 |

[a]Structures from the DUD set are in bold.

not expected to improve the accuracy, which is already exceptionally high. Adding other options (i.e, protein conformations) to the decoys led to some of them becoming false positives. With this specific protein, adding protein flexibility boosted the ranking of the decoys rather than improving the rankings of the actives. We are also aware of the limitations of the active/decoys set used herein that might be unintentionally biased toward a chemical series, which in turn prefers a given protein conformation.

All five PR structures led to AU-ROC at or below 0.70 (Table 3), however, used in conjunction with either flex or semiflex modes, the AU-ROC value was above 0.80 and warrants further discussion. In contrast to FGFr1, the flexible mode provides accuracy significantly better than any of the single structures. Our hypothesis was that some of the active ligands bind to any of the five structures, which are therefore all needed to dock all the actives properly.

To further investigate this hypothesis, the PR active ligands were clustered by similarity using our program SELECT (Figure 5) and the rankings of the actives were compared and analyzed (Figure 6). The clustering showed that half of the actives were clustered together while compounds 3, 16, and 19 could be part of this cluster if we lower the selection criterion (Figure 5). Another cluster made of six to eight compounds (depending on the selection criterion) appeared. Compound 6 was an outlier with Tanimoto coefficients below 0.70 with all the other compounds.

This clustering was then used to identify whether ligand diversity can be an important factor for docking-based VS on flexible proteins. When comparing similarity clusters and rankings, we observed that ranking of the compounds of the

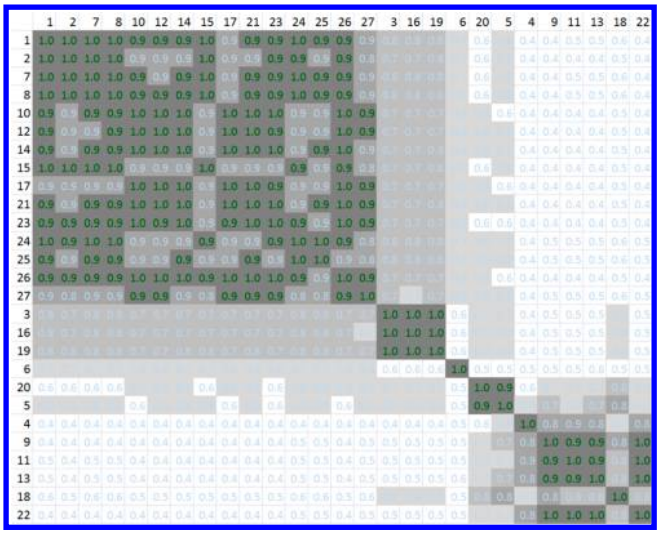


Figure 5. Clustering of the active PR inhibitors 1−27 by similarity (MACCS keys and Tanimoto coefficient) using SELECT.

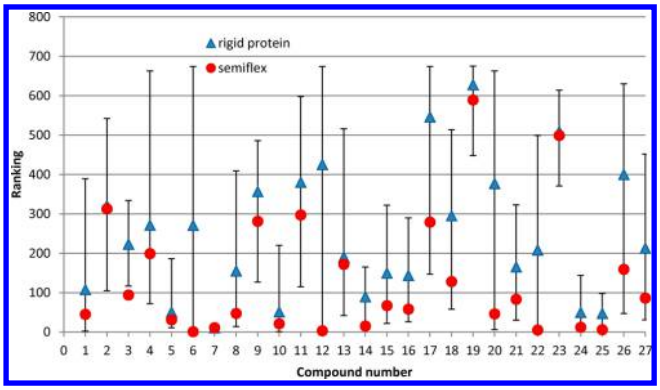



Figure 6. Rankings of the 27 active compounds when docked to rigid or semiflexible PR. The blue triangles are the average rankings over the five crystal structures, and the error bars represent the range of rankings over the five crystal structures.

first cluster were much more dependent on the protein structure used than compounds of the last cluster. More interestingly, for all the active compounds, the ranking in the semiflex mode was better than the median and average rankings in rigid mode. For example, closely related active compounds 2 and 14 (from the same cluster) were both ranked in the top two when 2w8y was used but ranked either in the top 10 (compound 14 with 1sr7 and 3d90) or over 250 (in all other cases) when the other protein structures were used. In semiflex mode, they were both found in the top three as if 2w8y only was used. The same observation was made with other compounds in other structures. For example, active 21 was found in the top 20 with 1sr7 and in semiflex mode and ranked lower with the other four structures in rigid mode. Overall, the ranking recorded when the protein was considered flexible was better in all cases than the average ranking when the compounds were docked to the rigid protein (Figure 6). This clearly demonstrated that flexible docking is able to accommodate protein structure for sets of dissimilar compounds.

We then expanded the study relating ligand diversity to the need for protein flexibility. For this purpose, the analysis of the improvement was made considering how diverse the known ligands are. As shown in Figure 7, when the actives are highly


dx.doi.org/10.1021/ci500299h | J. Chem. Inf. Model. 2014, 54, 3198−3210

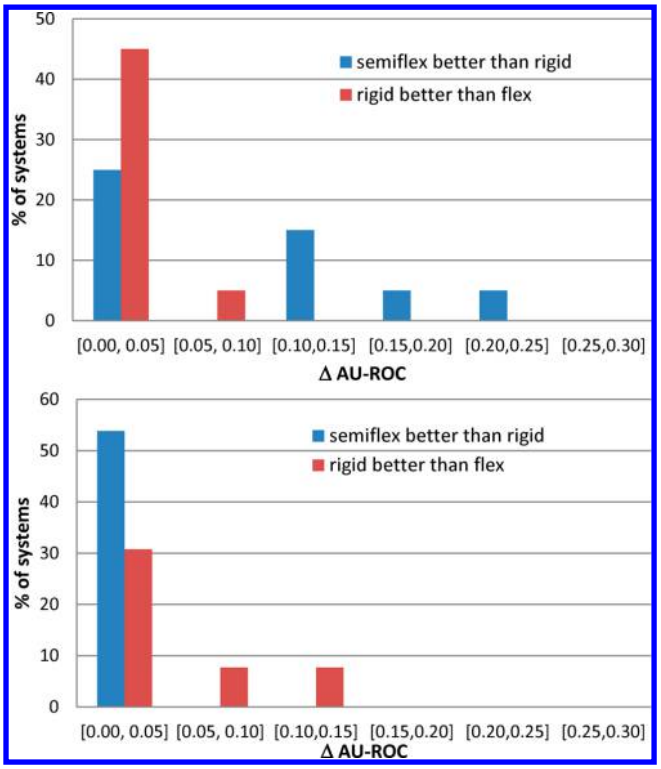



**Figure 7.** Ligand diversity and protein flexibility. (top) Average Tanimoto coefficient computed between all pairs of active compounds Tc < 0.70 (greater diversity); (bottom) Tc > 0.70 (lower diversity).

diverse (lower Tanimoto coefficient), the protein flexibility has overall a large and positive impact on the accuracy. In contrast, when the ligands are similar, considering protein flexibility is overall detrimental. This data clearly indicates that protein flexibility is required to accurately dock diverse sets of ligands.

Thus, our study revealed that without any prior knowledge on the preferred protein conformation, using the flexible protein mode semiflex would lead to an improvement of the AU-ROC greater than 0.05 in 33% of the cases and greater than 0.1 in 15% of the cases. In contrast, none of the systems showed a decrease greater than 0.1 and 12% a decrease greater than 0.05. Although, the balance is clearly in favor of the semiflex mode, selecting the protein flexibility mode should be done with care. The same analysis including the ligand diversity as a parameter was more instructive. Overall, the rigid mode is a right choice when all the screened chemical series are similar and are expected to bind to the same protein conformation. In this case, however, the right protein conformation (ie., the one the chemical series binds to) should be first identified either by crystallography with an analogue, or by comparative docking studies such as cross docking experiments. Docking to rigid proteins is therefore likely the right choice for lead optimization (docking of analogues). However, in a VS, it is unknown in advance which molecules will be active or to which conformation they will bind and making an informed selection of protein structures is hardly possible. As a result, using the flexible protein mode is likely to improve the accuracy of the screen by enabling the docking of structurally diverse compounds.

**Impact of the Presence of Water Molecules.** A number of research groups have developed methods such as WaterMap and 3D-RISM that accurately position and compute the free energy of binding of water molecules in protein binding site.[50,60−65] However, the impact of the inclusion of these key water molecules in proteins on the binding mode and binding affinity prediction has rarely been assessed. A study by Huang and Shoichet using 24 proteins from the DUD set (those including crystallographic water) showed that these water molecules can significantly improve the enrichment for half of the targets including COMT, FXa, AChE, AmpC, and NA.[66] WaterMap-derived WScore was also found to improve the enrichment when used with four proteins.[39] While this manuscript was in preparation, a report from Fukunishi and Nakamura demonstrated that the inclusion of the free energy of these water molecules significantly improved the scoring.[67]

In the current study, a thorough evaluation of the effect of the presence of water molecules in VS was completed. All but one of the proteins possess conserved water in at least one selected crystal structure. It is important to note that our implementations do not account for free energy of binding of water molecules and that only the impact of their occurrence is evaluated herein. The effects of keeping, discarding, and allowing water molecules or water particles to be displaceable when docking to rigid proteins are illustrated in Figures 8 and

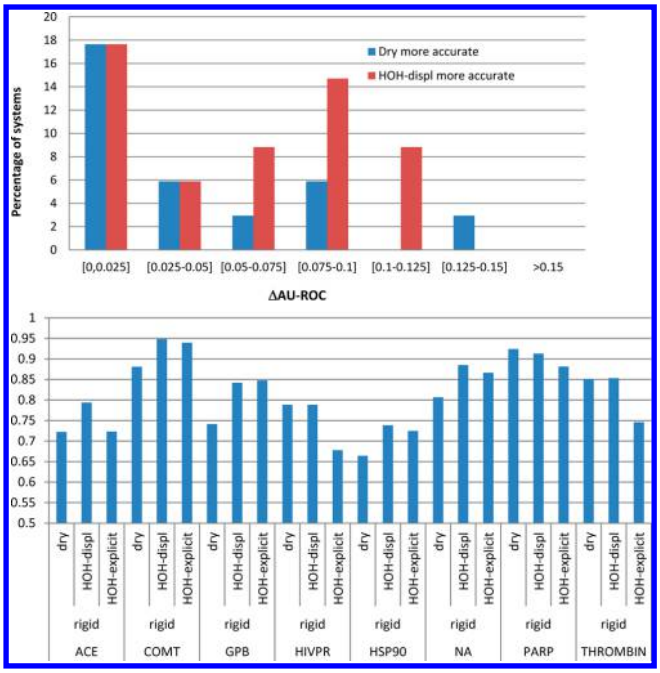



**Figure 8.** (top) AU-ROC with different water molecule implementations averaged over all the systems. (bottom) Example of significant effect of water molecules.

9. First, displaceable water improves slightly the results with most proteins. More interestingly, when explicit (non displaceable) water was used, the accuracy significantly increases or decreases in a very protein-dependent manner (Figure 8 bottom). When most of the active ligands require the presence of a given water molecule, keeping it is expected to improve the pose prediction accuracy (e.g., COMT, GPB). In contrast, when a given water molecule is to be displaced by several ligands, the use of explicit water molecules is expected to affect the accuracy (e.g., HIVPR, thrombin). In fact, a nondisplaceable water molecule should be carefully selected and the option of removing it when necessary must be available (as often suggested by the developers of other programs).[68]

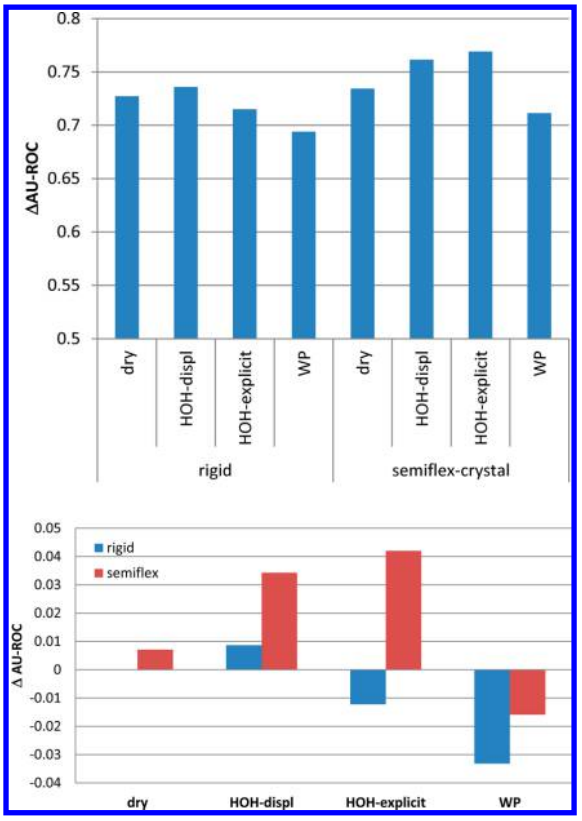



**Figure 9.** Impact of crystallographic water molecules or WPs in the overall VS accuracy compared to rigid-dry.

We mentioned above that the presence of displaceable water molecules did not significantly improve the docking pose prediction. A close look at the AU-ROC data revealed that the presence of explicit or displaceable water molecules led to higher accuracy than when water molecules were removed and protein flexibility was considered, while the accuracy was less affected when the docking experiments were carried out with single crystal structures (rigid mode). The use of WPs was much less successful with significant drops compared to crystallographic water molecules. This drop is likely the result of the poor placement of the water particles as discussed above.

In our implementation, the impact of water molecules on the VS accuracy is not as important as one would expect. We believe that the scoring of these key water molecules should be significantly improved. As shown by Friesner and co-workers, some water molecules have free energy of binding that can be much larger than others.[50] For instance, the presence of "crystalline" water molecules in the binding site of streptavidin can explain the exceptionally strong binding of biotin. As of now, our current scoring functions are not able to predict this contribution to the overall ligand/protein binding affinity. It is clear that the scoring of the individual explicit water molecules should be added to the state-of-the-art scoring functions.[16,63,69] In fact, in a protocol integrating WaterMap (WScore) and Glide showed significantly improved accuracy over Glide SP when applied to four enzymes.[39] Attempts to improve the accuracy include, for example, the consideration of the burying of interactions. We believe that this factor can be related to the burying of the water molecules before ligand binding hence can be a rough estimate of the free energy of the water that has been displaced upon ligand binding.[70]

**Fitted Accuracy—Protein Flexibility and Displaceable Water Molecules.** Overall, running a VS using a combination of protein flexibility and displaceable water molecules (AU-ROC = 0.76) is slightly better than docking to rigid protein using the protein structures selected by the DUD developers (AU-ROC = 0.74 and below, Figure 9) although this might not be statistically significant (no errors or intervals of confidence were systematically computed in this work). Reports on the evaluation of a number of other docking programs (i.e., ICM, Glide, Surflex, PhDock, DOCK, FlexX, and IMGDock) on the DUD sets identified Glide and Surflex as the most accurate programs with AU-ROC = 0.72 ± 0.05 and 0.66 ± 0.05 respectively.[31,71] More recently researchers from Schrödinger reported AU-ROC of 0.80 with Glide on the same DUD sets.[39] However, as we have shown throughout this manuscript, selecting other crystal structures might have provided improved accuracy as discussed above with PR. In fact, refined coordinates of the proteins provided improved accuracy with ICM with AU-ROC as high as 0.79[26] as well as with Surflex-Dock with AU-ROC as high as 0.83.[72]

**Remaining Challenges.** We previously observed that accounting for protein flexibility significantly improved the pose prediction over cross-docking. As questioned above, is protein flexibility improving the accuracy or adding noise to virtual screening campaigns? With this study, we have brought some indications that docking to flexible proteins generated improvement in the overall results. At this stage, analyzing the failures of the study can help improving the accuracy of docking in some more complex cases. We first found that in the case where all the actives are binding to the same conformation, introducing flexibility improved ranking of the decoys, but when the ligands are diverse enough and bind to several different conformations, introducing protein flexibility has a significantly positive effect. Thus, knowledge of the small molecule diversity may be a factor for making a decision on turning on or off protein flexibility when docking. To remedy this and other shortcomings, further improvement of the scoring function can be done by considering protein energy and water free energy of binding. Also, increasing the pose prediction accuracy will certainly improve the molecules scoring and ranking. To evaluate this last point we thought to compare the accuracy in pose prediction and the accuracy in active compound identification (Figure 10). Much to our surprise, we found no correlation between these two abilities of our docking program. Each protein shown in Figure 10 is represented by four or five crystal structures. This restricted number is not statistically high enough to enable the measure of the true pose prediction accuracy of our docking program for each protein and each set of 1000 actives and decoys.

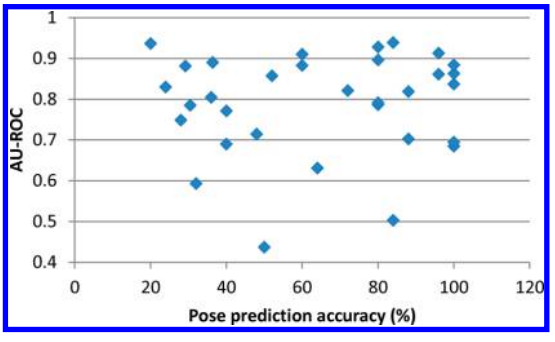



**Figure 10.** Virtual screening accuracy vs pose prediction accuracy.


dx.doi.org/10.1021/ci500299h | J. Chem. Inf. Model. 2014, 54, 3198−3210

## CONCLUSIONS

After examining the overall results from this VS study and specific examples, several reoccurring themes become clear. As expected, the accuracy of a VS will most likely always be protein and run-mode dependent (in addition to be program dependent). In some cases (such as ER), the run mode dependence will be minimal, (and the program dependence will be less pronounced). In other cases (such as PR), protein flexibility is critical. Running the docking experiments considering protein flexibility is overall advantageous. However, this is not categorically true, as was shown in the example of PPARg. The importance of the inclusion and use of waters is also an interesting point of discussion and is also very much protein dependent. We have found that FITTED generally produced better results with displaceable crystallographic water molecules and flexible proteins.

The overall conclusion from the collected data suggests that protein dependence and small molecule diversity-dependence can range from insignificant to highly significant, and therefore a high level of knowledge of both the library to be docked and of the protein must be used, when available prior to carrying out a VS. Unfortunately, when studying a new protein target, by definition, very little may be known about the protein and how key crystallographic waters or flexibility may affect binding. In these cases, the use of knowledge based on proteins with high amounts of homology with the protein in question might be used.

This study also indicates that in a lead optimization study (docking of analogous molecules), the rigid protein mode can be used as long as the protein has been cocrystallized with a structurally similar compound. In contrast, in a large VS campaign looking for new chemical series, the protein flexibility should be considered.

This study also suggests that significant changes and/or additions to the existing scoring functions are required to account for the free energy of the water (kept or displaced) and the free energy of the protein, two factors being currently investigated.

## EXPERIMENTAL SECTION

**Docking with FITTED.** The subversion 3801 of the FORECASTER platform including all the necessary programs used in this work (FITTED, CONVERT, SMART, SELECT, PREPARE, PROCESS, MATCH-UP) has been used for this study.

**Construction of the Testing Sets.** The sets of decoys and ligands were downloaded from the DUD Web site and further processed as follows. First the hydrogens were removed from all the molecules in order to convert the tautomers into identical molecules using a routine of FORECASTER and to reassign the bond order. Then the molecules were clustered by similarity using SELECT and 40 ligands and 960 decoys with the largest diversity were selected whenever available or less but with the same ligands/decoys ratio otherwise. Hydrogens were added using CONVERT and all these ligands prepared for docking using SMART.

**Preparation of the Protein Files.** MATCH-UP, PREPARE, and PROCESS were applied with the default parameters and the specific keyword identifying metalloenzymes when applicable (ACE, ADA, and PDE5).

**Docking with FITTED.** Default parameters implemented in FITTED have been used.

## ASSOCIATED CONTENT

### Supporting Information

All the AU-ROC collected for all the settings and proteins. This material is available free of charge via the Internet at http://pubs.acs.org. The program is available on request from the authors (www.fitted.ca).

## AUTHOR INFORMATION

### Corresponding Author

*Phone: 514-398-8543. Fax: 514-398-3797. E-mail: nicolas.moitessier@mcgill.ca.

### Notes

The authors declare the following competing financial interest(s): The program FITTED used in this study is distributed (free to academia) by Molecular Forecaster co-founded by N.M. and E.T.

## ACKNOWLEDGMENTS

We thank NSERC (Discovery program) for financial support. Calcul Québec and Compute Canada are acknowledged for generous CPU allocations. E.T. was supported by a Fessenden Professorship to N.M., N.W. was supported by a fellowship from GRASP (FRSQ).

## REFERENCES

(1) Schulz, M. N.; Hubbard, R. E. Recent progress in fragment-based lead discovery. *Curr. Opin. Pharmacol.* **2009**, *9*, 615−621.

(2) Clark, D. E. What has virtual screening ever done for drug discovery? *Exp. Opin. Drug Discovery* **2008**, *3*, 841−851.

(3) Koppen, H. Virtual screening - what does it give us? *Curr. Opin. Drug Discovery Devel.* **2009**, *12*, 397−407.

(4) McInnes, C. Virtual screening strategies in drug discovery. *Curr. Opin. Chem. Biol.* **2007**, *11*, 494−502.

(5) Zoete, V.; Grosdidier, A.; Michieli, O. Docking, virtual high throughput screening and in silico fragment-based drug design. *J. Cell Mol. Med.* **2009**, *13*, 238−248.

(6) Ferreira, R. S.; Simeonov, A.; Jadhav, A.; Eidam, O.; Mott, B. T.; Keiser, M. J.; McKerrow, J. H.; Maloney, D. J.; Irwin, J. J.; Shoichet, B. K. Complementarity Between a Docking and a High-Throughput Screen in Discovering New Cruzain Inhibitors. *J. Med. Chem.* **2010**, *53*, 4891−4905.

(7) Therrien, E.; Englebienne, P.; Arrowsmith, A. G.; Mendoza-Sanchez, R.; Corbeil, C. R.; Weill, N.; Campagna-Slater, V.; Moitessier, N. Integrating medicinal chemistry, organic/combinatorial chemistry, and computational chemistry for the discovery of selective estrogen receptor modulators with FORECASTER, a novel platform for drug discovery. *J. Chem. Inf. Model.* **2012**, *52*, 210−224.

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

(9) Andricopulo, A. D.; Salum, L. B.; Abraham, D. J. Structure-based drug design strategies in medicinal chemistry. *Curr. Top. Med. Chem.* **2009**, *9*, 771−790.

(10) Corbeil, C. R.; Therrien, E.; Moitessier, N. Modeling reality for optimal docking of small molecules to biological targets. *Curr. Comput.-Aided Drug Des.* **2009**, *5*, 241−263.

(11) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br. J. Pharmacol.* **2008**, *153*, S7−S26.

(12) May, A.; Zacharias, M. Protein-ligand docking accounting for receptor side chain and global flexibility in normal modes: Evaluation on kinase inhibitor cross docking. *J. Med. Chem.* **2008**, *51*, 3499−3506.

(13) Jain, A. Effects of protein conformation in docking: improved pose prediction through protein pocket adaptation. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 355−374.

(14) Claußen, H.; Buning, C.; Rarey, M.; Lengauer, T. FLEXE: Efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308*, 377−395.
(15) García-Sosa, A. T. Hydration Properties of Ligands and Drugs in Protein Binding Sites: Tightly-Bound, Bridging Water Molecules and Their Effects and Consequences on Molecular Design Strategies. *J. Chem. Inf. Model.* **2013**, *53*, 1388−1405.
(16) De Beer, S. B. A.; Vermeulen, N. P. E.; Oostenbrink, C. The role of water molecules in computational drug design. *Curr. Top. Med. Chem.* **2010**, *10*, 55−66.
(17) Rarey, M.; Kramer, B.; Lengauer, T. The particle concept: Placing discrete water molecules during protein- ligand docking predictions. *Prot. Struct. Funct. Genet.* **1999**, *34*, 17−28.
(18) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem.* **2005**, *48*, 6504−6515.
(19) Forli, S.; Olson, A. J. A force field with discrete displaceable waters and desolvation entropy for hydrated ligand docking. *J. Med. Chem.* **2012**, *55*, 623−638.
(20) Morris, G. M.; Ruth, H.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785−2791.
(21) *FlexX*, Release 2; BioSolveIT: Sankt Augustin, Germany, 2014.
(22) *GOLD*, 3.0; CCDC: Cambridge, UK, 2005.
(23) Korb, O.; Stützle, T.; Exner, T. An ant colony optimization approach to flexible protein−ligand docking. *Swarm Intel.* **2007**, *1*, 115−134.
(24) Korb, O.; Stutzle, T.; Exner, T. E. Empirical Scoring Functions for Advanced Protein/Ligand Docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49*, 84−96.
(25) Cavasotto, C. N.; Abagyan, R. A. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *J. Mol. Biol.* **2004**, *337*, 209−225.
(26) Neves, M. C.; Totrov, M.; Abagyan, R. Docking and scoring with ICM: the benchmarking results and strategies for improvement. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 675−686.
(27) Elokely, K. M.; Doerksen, R. J. Docking Challenge: Protein Sampling and Molecular Docking Performance. *J. Chem. Inf. Model.* **2013**, *53*, 1934−1945.
(28) Rabal, O.; Schneider, G.; Borrell, J. I.; Teixido, J. Structure-based virtual screening of FGFR inhibitors: Cross-decoys and induced-fit effect. *BioDrugs* **2007**, *21*, 31−45.
(29) Corbeil, C. R.; Moitessier, N. Docking Ligands into Flexible and Solvated Macromolecules. 3. Impact of Input Ligand Conformation, Protein Flexibility, and Water Molecules on the Accuracy of Docking Programs. *J. Chem. Inf. Model.* **2009**, *49*, 997−1009.
(30) Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 4. Are popular scoring functions accurate for this class of proteins? *J. Chem. Inf. Model.* **2009**, *49*, 1568−1580.
(31) Schumann, M.; Armen, R. S. Systematic and efficient side chain optimization for molecular docking using a cheapest-path procedure. *J. Comput. Chem.* **2013**, *34*, 1258−1269.
(32) Corbeil, C. R.; Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *J. Chem. Inf. Model.* **2007**, *47*, 435−449.
(33) Corbeil, C. R.; Englebienne, P.; Yannopoulos, C. G.; Chan, L.; Das, S. K.; Bilimoria, D.; Heureux, L.; Moitessier, N. Docking Ligands into Flexible and Solvated Macromolecules. 2. Development and Application of FITTED 1.5 to the Virtual Screening of Potential HCV Polymerase Inhibitors. *J. Chem. Inf. Model.* **2008**, *48*, 902−909.
(34) Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 5. Force-field-based prediction of binding affinities of ligands to proteins. *J. Chem. Inf. Model.* **2009**, *49*, 2564−2571.
(35) Pottel, J.; Therrien, E.; Gleason, J. L.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 6. Development and application to the docking of HDACs and other zinc metalloenzymes inhibitors. *J. Chem. Inf. Model.* **2014**, *54*, 254−265.
(36) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789−6801.
(37) Schneider, N.; Hindle, S.; Lange, G.; Klein, R.; Albrecht, J.; Briem, H.; Beyer, K.; Claußen, H.; Gastreich, M.; Lemmen, C.; Rarey, M. Substantial improvements in large-scale redocking and screening using the novel HYDE scoring function. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 701−723.
(38) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55*, 6582−6594.
(39) Repasky, M.; Murphy, R.; Banks, J.; Greenwood, J.; Tubert-Brohman, I.; Bhat, S.; Friesner, R. Docking performance of the glide program as evaluated on the Astex and DUD datasets: a complete set of glide SP results and selected results for a new scoring function integrating WaterMap and glide. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 787−799.
(40) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.
(41) Najmanovich, R.; Kuttner, J.; Sobolev, V.; Edelman, M. Side-chain flexibility in proteins upon ligand binding. *Prot. Struct. Funct. Genet.* **2000**, *39*, 261−268.
(42) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The penultimate rotamer library. *Prot. Struct. Funct. Genet.* **2000**, *40*, 389−408.
(43) Subramaniam, S.; Senes, A. An energy-based conformer library for side chain optimization: Improved prediction and adjustable sampling. *Prot. Struct. Funct. Bioinf.* **2012**, *80*, 2218−2234.
(44) De Groot, B. L.; Van Aalten, D. M. F.; Scheek, R. M.; Amadei, A.; Vriend, G.; Berendsen, H. J. C. Prediction of protein conformational freedom from distance constraints. *Prot. Struct. Funct. Genet.* **1997**, *29*, 240−251.
(45) Seeliger, D.; Haas, J.; de Groot, B. L. Geometry-Based Sampling of Conformational Transitions in Proteins. *Structure* **2007**, *15*, 1482−1492.
(46) Sims, G. E.; Kim, S. H. A method for evaluating the structural quality of protein models by using higher-order $\varphi$-$\psi$ pairs scoring. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4428−4432.
(47) Sims, G. E.; Choi, I. G.; Kim, S. H. Protein conformational in higher order $\varphi$-$\psi$ maps. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 618−621.
(48) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(49) Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Effects of Water Placement on Predictions of Binding Affinities for p38$\alpha$ MAP Kinase Inhibitors. *J. Chem. Theory Comput.* **2010**, *6*, 3850−3856.
(50) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 808−813.
(51) Truchon, J. F.; Bayly, C. I. Evaluating virtual screening methods: Good and bad metrics for the "early recognition" problem. *J. Chem. Inf. Model.* **2007**, *47*, 488−508.
(52) Clark, R. D.; Webster-Clark, D. J. Managing bias in ROC curves. *J. Comp. Aided Mol. Des.* **2008**, *22*, 141−146.
(53) Zweig, M. H.; Campbell, G. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clin. Chem.* **1993**, *39*, 561−577.
(54) Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes? *J. Comp. Aided Mol. Des.* **2008**, *22*, 213−228.
(55) Jain, A. N. Bias, reporting, and sharing: Computational evaluations of docking methods. *J. Comp. Aided Mol. Des.* **2008**, *22*, 201−212.

(56) Jain, A. N.; Nicholls, A. Recommendations for evaluation of computational methods. *J. Comp. Aided Mol. Des.* **2008**, *22*, 133−139.
(57) Bauer, M. R.; Ibrahim, T. M.; Vogel, S. M.; Boeckler, F. M. Evaluation and optimization of virtual screening workflows with DEKOIS 2.0 - A public library of challenging docking benchmark sets. *J. Chem. Inf. Model.* **2013**, *53*, 1447−1462.
(58) Liebeschuetz, J.; Cole, J.; Korb, O. Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. *J. Comp. Aided Mol. Des.* **2012**, *26*, 737−748.
(59) Shan, Y.; Arkhipov, A.; Kim, E. T.; Pan, A. C.; Shawa, D. E. Transitions to catalytically inactive conformations in EGFR kinase. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 7270−7275.
(60) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of the water content in protein binding sites. *J. Phys. Chem. B* **2009**, *113*, 13337−13346.
(61) Zheng, M.; Li, Y.; Xiong, B.; Jiang, H.; Shen, J. Water PMF for predicting the properties of water molecules in protein binding site. *J. Comput. Chem.* **2013**, *34*, 583−592.
(62) Wang, L.; Berne, B. J.; Friesner, R. A. Ligand binding to protein-binding pockets with wet and dry regions. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 1326−1330.
(63) Nikolić, D.; Blinov, N.; Wishart, D.; Kovalenko, A. 3D-RISM-Dock: A new fragment-based drug design protocol. *J. Chem. Theory Comput.* **2012**, *8*, 3356−3372.
(64) Kiyota, Y.; Yoshida, N.; Hirata, F. A new approach for investigating the molecular recognition of protein: Toward structure-based drug design based on the 3D-RISM theory. *J. Chem. Theory Comput.* **2011**, *7*, 3803−3815.
(65) Nguyen, C. N.; Cruz, A.; Gilson, M. K.; Kurtzman, T. Thermodynamics of Water in an Enzyme Active Site: Grid-Based Hydration Analysis of Coagulation Factor Xa. *J. Chem. Theory Comput.* **2014**, *10*, 2769−2780.
(66) Huang, N.; Shoichet, B. K. Exploiting Ordered Waters in Molecular Docking. *J. Med. Chem.* **2008**, *51*, 4862−4865.
(67) Fukunishi, Y.; Nakamura, H. Improved estimation of protein-ligand binding free energy by using the ligand-entropy and mobility of water molecules. *Pharmaceuticals* **2013**, *6*, 604−622.
(68) http://www.schrodinger.com/kb/31 (accessed Oct. 24, 2014).
(69) Guimares, C. R. W.; Mathiowetz, A. M. Addressing limitations with the MM-GB/SA scoring procedure using the watermap method and free energy perturbation calculations. *J. Chem. Inf. Model.* **2010**, *50*, 547−559.
(70) O'Boyle, N. M.; Brewerton, S. C.; Taylor, R. Using buriedness to improve discrimination between actives and inactives in docking. *J. Chem. Inf. Model.* **2008**, *48*, 1269−1278.
(71) Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y. B.; Humblet, C. Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2009**, *49*, 1455−1474.
(72) Spitzer, R.; Jain, A. Surflex-Dock: Docking benchmarks and real-world application. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 687−699.