JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Docking Challenge: Protein Sampling and Molecular Docking Performance

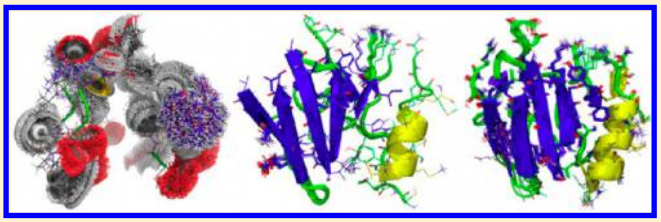
Khaled M. Elokely[†] and Robert J. Doerksen*[,†,‡]

[†]Department of Medicinal Chemistry, School of Pharmacy, University of Mississippi, University, Mississippi 38677, United States

[‡]Research Institute of Pharmaceutical Sciences, School of Pharmacy, University of Mississippi, University, Mississippi 38677, United States

Supporting Information

ABSTRACT: Computational tools are essential in the drug design process, especially in order to take advantage of the increasing numbers of solved X-ray and NMR protein−ligand structures. Nowadays, molecular docking methods are routinely used for prediction of protein−ligand interactions and to aid in selecting potent molecules as a part of virtual screening of large databases. The improvements and advances in computational capacity in the past decade have allowed for further developments in molecular docking algorithms to address more complicated aspects such as protein flexibility. The effects of incorporation of active site water molecules and implicit or explicit solvation of the binding site are other relevant issues to be addressed in the docking procedures. Using the right docking algorithm at the right stage of virtual screening is most important. We report a staged study to address the effects of various aspects of protein flexibility and inclusion of active site water molecules on docking effectiveness to retrieve (and to be able to predict) correct ligand poses and to rank docked ligands in relation to their biological activity for CHK1, ERK2, LpxC, and UPA. We generated multiple conformers for the ligand and compared different docking algorithms that use a variety of approaches to protein flexibility, including rigid receptor, soft receptor, flexible side chains, induced fit, and multiple structure algorithms. Docking accuracy varied from 1% to 84%, demonstrating that the choice of method is important.



1. INTRODUCTION

Protein−ligand docking is a powerful tool to study and provide a proper understanding of protein−ligand interactions. Docking is regularly used in different stages of drug design strategies, such as to facilitate design of potentially active leads.[1,2] Detection of the best ligand poses and proper ranking of several ligands' relative docking propensity are of great importance. Molecular docking, in practice, has two essential requirements:[3] structural data, for candidate ligands and the protein target of interest and a procedure to estimate protein−ligand interaction poses and strengths.[4] The RSCB Protein Data Bank (PDB) repository[5,6] is the main source of protein target structures for docking studies. The number of structures deposited in the PDB repository has been rapidly increasing for many years. Currently there are >62,000 PDB entries of protein−ligand complexes, of which >60,000 were solved by X-ray and >1700 by NMR methods (other techniques were used to solve the remaining structures).[7] The candidate ligands in docking procedures are generally small molecules. There was a rapid increase in the number of available synthesized chemical libraries after the development of combinatorial chemistry,[8] which increased demand for the development of fast and cheap ways to test interactions with protein targets. The increasing numbers of PDB entries and of chemical database entries, coupled to the strong desire to be able to predict binding modes and binding affinities of ligands, has led to a wide acceptance of the routine use of docking methods as a crucial step in virtual screening.[9] Various molecular docking algorithms are available to predict protein−ligand poses and to rank them based on scoring functions implemented in each specific docking approach.[10,11] Practically, docking software applications require protein−ligand sampling algorithms in order to be able to generate acceptable ligand poses. Ligand sampling algorithms, for ligand pose generation and placement in the active site, are of three types: shape matching,[12,13] systematic search,[14] and stochastic algorithms.[15] Ligand conformational sampling is an essential step that generates a ligand multiconformer database to be used in ligand sampling. Conformational search is sometimes performed as a separate step before docking[16] or can be implemented as an integrated part of the docking algorithm.[17] Protein sampling refers to the allowed degree of binding site flexibility. Docking algorithms may consider the protein as a rigid body,[12,18] as a soft body,[14,19,20] to have flexible side chains,[19−21] or to have certain flexible domains.[22−24] Alternatively, protein flexibility can be represented by using multiple conformers or ensembles of rigid protein structures.[25] Various classes of scoring functions are

Special Issue: 2012 CSAR Benchmark Exercise

Received: January 19, 2013
Published: March 26, 2013

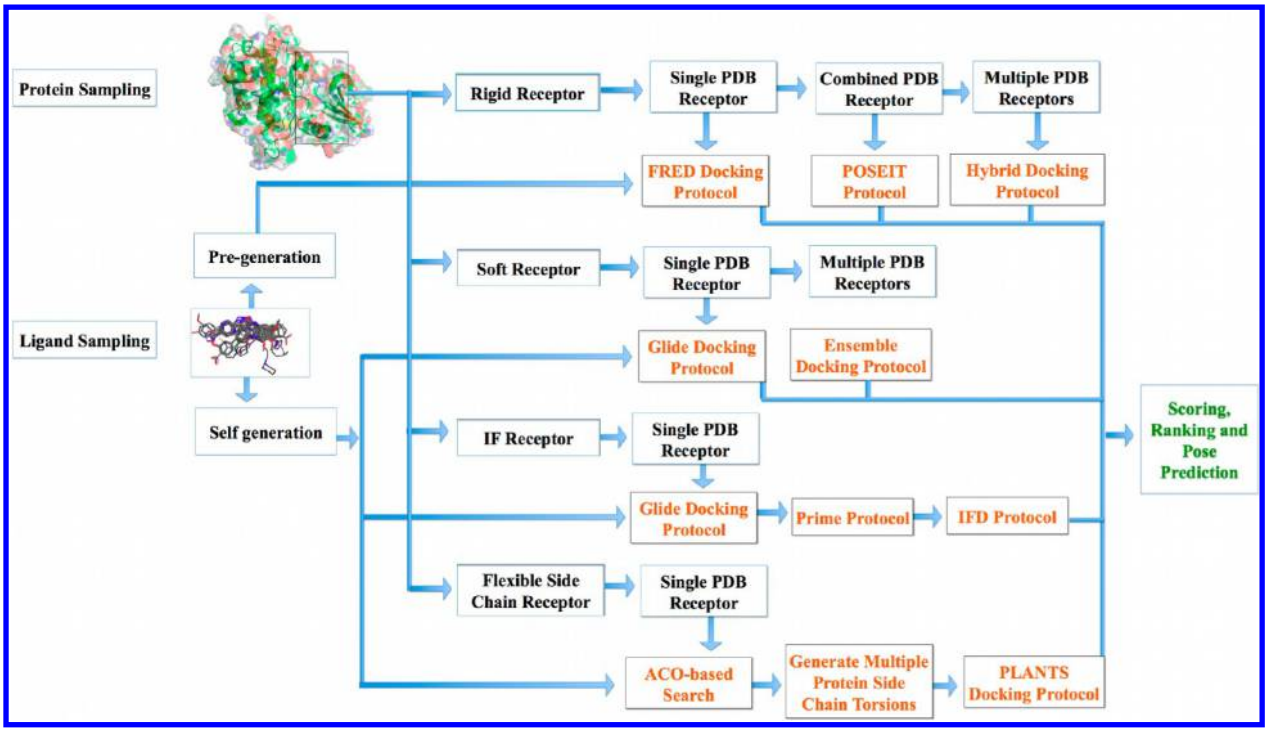


Figure 1. Study protocol starts with ligand and protein sampling, followed by setting up docking calculations with subsequent scoring, ranking, and pose prediction of the docked ligands. Protein sampling accounts for five approaches: rigid receptor, soft receptor, multiple (ensemble) receptors, FSC receptor, and IF receptor. Ligand sampling is either precedent to the docking procedure or a part of the specific docking technique. Scoring, ranking, and pose prediction are carried out in relation to known active and co-crystallized ligands.

used to estimate the binding affinities of ligand poses.[26] Scoring functions can be classified as force field-based,[27,28] empirical,[29,30] knowledge-based,[31,32] clustering and entropy-based,[33−35] or consensus scoring methods.[36−38] Active site water molecules can be considered another aspect of docking target flexibility.[39,40] Incorporation of active site water molecules in the docking procedure is challenging. Each water molecule needs to be analyzed to check if it is an integral part of the protein or just an artifact of the crystallization procedure.

In this work, which was part of the annual Community Structure−Activity Resource (CSAR) challenge, we studied the ability of different protein−ligand molecular docking algorithms to regenerate the correct ligand binding mode of crystal structure bound ligands and to rank active ligands with respect to their activity data. The study involved a five-stage docking approach based on the degree of allowed protein flexibility (Figure 1).

2. METHODS

2.1. Ligand Databases Collection. Ligand databases for all target proteins in the study were downloaded from CSAR. Initial databases did not include the activity data, whereas the final databases did include it. Protein-bound ligands were included in both the initial and final databases to serve as a check on pose prediction accuracy.

2.1.1. Serine/Threonine Protein Kinase Chk1 (CHK1). The structure of CHK1 (also called checkpoint kinase 1) was downloaded from the RCSB PDB repository (PDB ID: 2E9N)[41] and used for primary study. A database of 17 additional PDB structures was downloaded in mol2 format from CSAR[42] and in pdb format from the RCSB PDB (PDB IDs: 4FSM, 4FSN, 4FSQ, 4FSR, 4FST, 4FSU, 4FSW, 4FSY, 4FSZ, 4FT0, 4GH2, 4FT3, 4FT5, 4FT7, 4FT9, 4FTA, and 4FTC).

A database of 47 ligand structures was obtained from CSAR to be used in the primary study. Subsequently a final database of 184 ligands including the previous 47 was obtained from CSAR.

2.1.2. Extracellular Signal-Regulated Kinase 2 (ERK2). For primary study of ERK2 (also called mitogen-activated kinase 1 or MAPK1), we downloaded 3I5Z.pdb[43] from the RCSB PDB repository. For secondary study, we downloaded 12 protein mol2 structures from CSAR and the same ones in pdb format from the RCSB PDB repository (PDB IDs: 4FUX, 4FUY, 4FV0, 4FV1, 4FV2, 4FV3, 4FV4, 4FV5, 4FV6, 4FV7, 4FV8, and 4FV9). The initial ligand database contained 39 structures, and the final database was extended to include a total of 52 structures.

2.1.3. UDP-3-O-N-Acetylglucosamine Deacetylase (LpxC) of Pseudomonas aeruginosa. The PDB structure of LpxC (PDB ID: 3P3E)[44] was downloaded for the primary study. We downloaded five other protein structures in mol2 format from CSAR, four of which were deposited in pdb format. The PDB structures have been deposited in the RSCB PDB repository (PDB IDs: 4FW3, 4FW4, 4FW5, 4FW6, and 4FW7) but are not yet released. The initial ligand database consisted of 16 ligands, whereas the final database was extended to a total of 31 ligands.

2.1.4. Urokinase-Type Plasminogen Activator (UPA). We retrieved the pdb structure of UPA (also known simply as urokinase or urokinase plasminogen activator) from the PDB (PDB ID: 1OWE).[45] Seven UPA structures were downloaded from CSAR as mol2 files and from the RSCB PDB repository as pdb files (PDB IDs: 4FU7, 4FU8, 4FU9, 4FUB, 4FUC, 4FUD, and 4FUE). The initial ligand database consisted of 20

structures, but that was extended to 46 structures in the final database.

2.2. Protein Preparation with Protein Preparation Wizard.[46] The PDB protein−ligand structures were processed with the Protein Preparation Wizard in the Schrödinger suite.[47] The protein structure integrity was checked and adjusted, and missing residues and loop segments near the active site were added using Prime.[48−50] Hydrogen atoms were added after deleting any original ones, followed by adjustment of bond orders for amino acid residues and the ligand. The protonation and tautomeric states of Asp, Glu, Arg, Lys, and His were adjusted to match a pH of 7.4. Possible orientations of Asn and Gln residues were generated. Active site water molecules beyond 5.0 Å from the ligand were deleted. Hydrogen bond sampling with adjustment of active site water molecule orientations was performed using PROPKA[51] (propka.ki.ku.dk) at pH 7.4. Water molecules with fewer than two hydrogen bonds to non-waters were deleted. Then, the protein−ligand complex was subjected to geometry refinement using an OPLS2005 force field[52] restrained minimization with convergence of heavy atoms to an RMSD of 0.3 Å.

2.3. Ligand Preparation with Ligprep.[53] Ligands were prepared using Ligprep from the Schrödinger suite. We obtained the initial ligand databases as collections of SMILES (simplified molecular-input line-entry system) strings (which do not contain 3D coordinates). The final ligand databases were in the mol2 format (3D structures). We included all structures without performing predocking filtering. We generated a single low energy 3D conformer with acceptable bond lengths and angles for each 2D structure in the initial databases. For the initial and final databases, after 3D structure generation, we prepared ligand structures for molecular docking. Ligprep used the OPLS2005 force field and charges in all ligand preparation steps. All possible protomers (protonation states) and ionization states were enumerated for each ligand using Ionizer at a pH of 7.4. Stereoisomers were generated for the five structures with unassigned stereogenic centers, with a limit of 32 stereoisomers considered per ligand. Tautomeric states were generated for chemical groups with possible prototropic tautomerism. Only the lowest energy conformer was kept for each ligand.

2.4. Ligand Conformational Sampling. Ligand conformational sampling was used with the FRED[54] and HYBRID[55] modules of OpenEye. Other molecular docking software applications we used in this study have their own integrated conformational sampling algorithms. We used OMEGA 2.4.6 (OpenEye)[56,57] as the conformer generator. The SMILES notations of the initial databases and mol2 structures of the final databases were used as input for OMEGA 2.4.6. The conformational search force field was defined as the 94s variant of the Merck Molecular Force Field (MMFF94s).[58] We kept all generated conformers within a 10.0 kcal/mol energy window, except that to eliminate redundant conformers an RMSD cutoff of 0.5 Å was used.

2.5. Protein Sampling. To address protein flexibility in molecular docking, a large number of degrees of freedom should be considered. We performed a staged study starting from rigid body docking and continuing all the way to fully flexible active site docking.

2.5.1. Docking Using Rigid Receptors. The OEDocking v3.0.0 distribution using the FRED ligand shape fitting algorithm was utilized for receptor rigid body docking. All receptors used in this study were co-crystallized with ligands. The bound ligands were used to specify the active site. A 3D box was generated around each ligand to enclose the active site. Because we do not have any extremely large active site in our study, a negative image potential was created for each active site with disabled inner contour. No constraints were added except for the LpxC target, for which we prepared receptor docking sites with and without a Zn^{2+} metal constraint. We prepared each receptor with and without active site water molecules. We saved the multiconformer ligand files in OEBinary, and therefore, there was no need to use the FRED conformer test flag. FRED was used with standard docking precision using 1.0 Å for the ligand translational step size and 1.5 Å for the rotational step size. Because the databases we used are small, we maintained the default value of keeping the 500 top scoring molecules with a maximum of one pose to be saved for each molecule.

2.5.2. Docking Using Soft Receptors. We used Glide 5.8[59] for soft receptor molecular docking. The receptor grid for each target was prepared using the OPLS2005 force field. We specified the area surrounding the co-crystallized ligand as the receptor binding pocket. We excluded all bound ligands from the grid generation, except for Zn^{2+} in the case of LpxC and active site water molecules when applicable. Softening the potential of the nonpolar parts of the receptor was carried out by scaling the van der Waals radii by a factor of 0.8. Atoms were considered as nonpolar if their absolute partial atomic charge was determined to be $\leq$0.25. The grid center was set to be the centroid of the co-crystallized ligand, and the cubic grid had a side length of 10 Å. No constraints were used in any of the receptor grids. Rotations were allowed for the hydroxyl groups in Ser, Thr, and Tyr, and the thiol group in Cys residues. After grid preparation, prepared ligand databases were docked into the generated receptor grids using Glide SP docking precision. Flexible ligand sampling was considered in the docking procedure. For the docking runs, a second softening potential was considered. A 0.8 scaling factor was used for van der Waals radii of the ligands' nonpolar atoms with absolute partial atomic charge $\leq$0.15. All poses were subjected to post-docking minimization. We saved the best-docked structure for each ligand, based on the model energy score which combines the energy grid score, the binding affinity, the internal strain energy, and the Coulomb−van der Waals interaction energy scores.

2.5.3. Docking Using Receptors with Flexible Side Chains. We used the protein−ligand ant system (PLANTS)[60] to deal with side chain flexibility of amino acid residues. PLANTS uses the artificial ant colony optimization (ACO) algorithm to find the best ligand pose in the binding pocket. ZODIAC 0.65[61] was used to prepare PLANTS input files. All ligands and protein structures were preprocessed by the structure protonation and recognition system (SPORES)[62] to adjust the protonation and tautomeric states and to assign stereoisomers for non-specified asymmetric centers. The binding site was specified from the co-crystallized ligand coordinates. We used normal ligand sampling search speed with 20 ants and simplex rescoring. We used values for the sigma parameter (σ = 1.0) and evaporation rate (ρ = 0.5) that have been shown to be sufficient with 20 ants.[63] Planar bond rotations were forced. For clustering, an RMSD of 2 Å and a maximum of 10 clusters were defined. CHEMPLP[64] was specified as the scoring function. All amino acids in the defined binding site were selected to have flexible side chains during docking.

2.5.4. Docking Using Flexible Binding Domain for Receptors, or Induced Fit Docking (IFD). Ligands are known

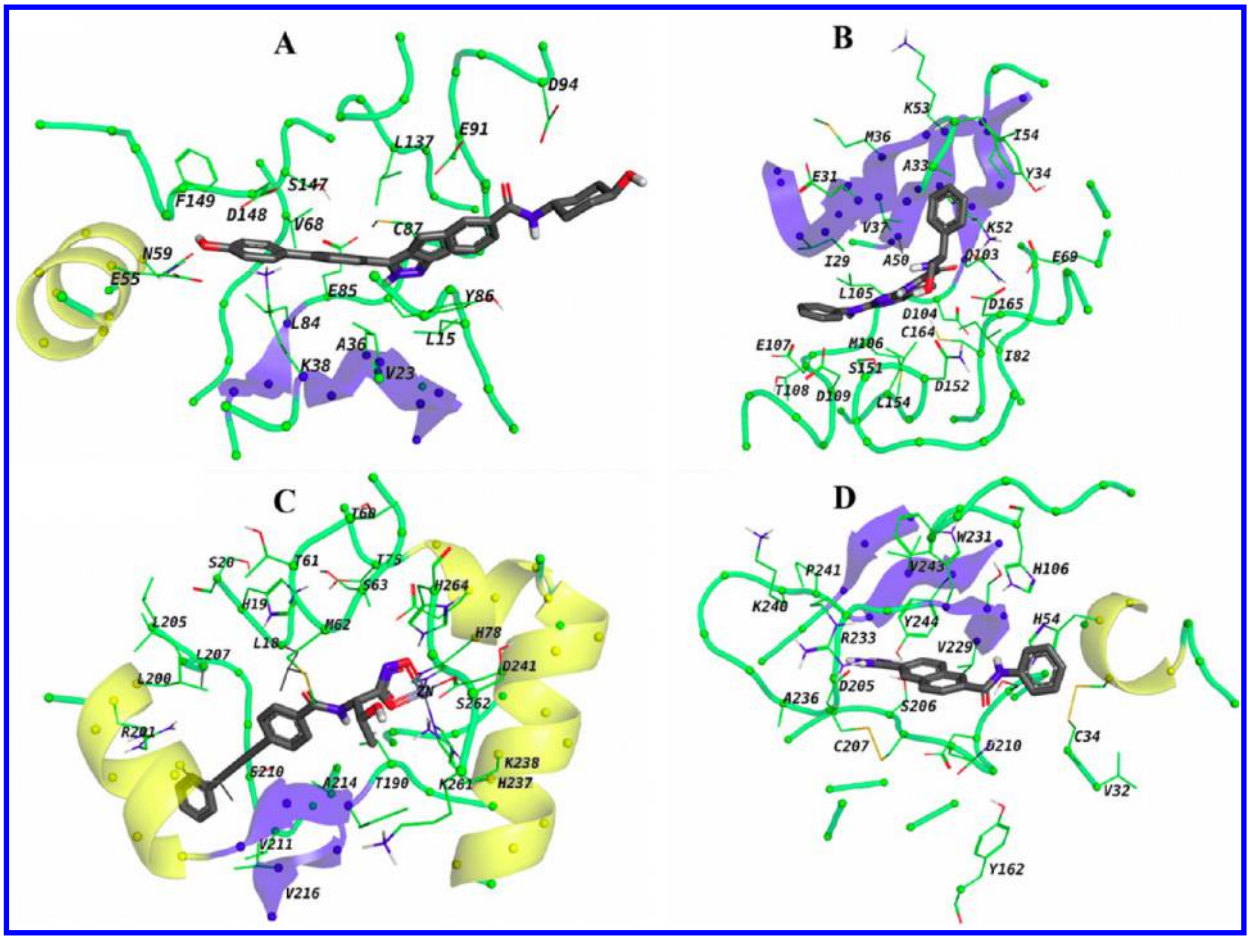


Figure 2. Active site structures of target proteins; CHK1 (PDB ID: 2E9N, panel A), ERK2 (PDB ID: 3I5Z, panel B), LpxC (PDB ID: 3P3E, panel C) and UPA (PDB ID: 1OWE, panel D). The images were generated using PyMol.[72] The α-helices, β-sheets, and loops are colored in yellow, blue, and green, respectively. Key amino acids in the active sites have their side chains displayed as lines and are labeled based on their position. Only polar hydrogen atoms are displayed in white. All Cα atoms are represented as spheres and colored according to the corresponding secondary structure. Other carbons of the amino acids are colored green. Ligands are displayed as sticks with gray carbons, and no hydrogen atoms are depicted. Oxygen, nitrogen, and sulfur atoms are colored red, blue, and yellow, respectively.

often to induce conformational changes in the active site upon binding. We used the Schrödinger induced fit docking (IFD) protocol to represent this. The receptor grid center was specified from the bound ligand, and the cubic grid had a side length of 10 Å. A 2.5 kcal/mol energy window that was used for ligand conformational sampling. The scaling factors to soften the potentials of the receptors and ligands were set to 0.5 in both cases. A maximum of 20 poses was saved. All residues within 5.0 Å of ligand poses were refined using the Prime molecular dynamics module to allow for binding domain flexibility. The Schrödinger IFD software allows specification of whole loops to move, but we found that if the key linking residues are flexible then the whole loop can adjust to fold differently without needing to use that option. Glide SP was used for the redocking step into the top 20 receptor structures generated within 30 kcal/mol of the best structure by the Prime refinement.

2.5.5. Docking Using an Ensemble of Protein Structures. Another way to address protein flexibility is to use multiple or an ensemble of protein structures. OEDocking v3.0.0 with its HYBRID program was used for this. We used the same options as in FRED docking but with multiple structures of each target. All protein structures were treated as rigid.

The Schrödinger suite has the same capability of ensemble docking but with a soft receptor approach. Receptor grids were prepared as described above. The van der Waals scaling factor was specified as 0.8 for receptor nonpolar atoms, and a partial charge cutoff of 0.15 was used. Glide SP was used for docking, and one pose was saved for each ligand.

2.6. Pose Prediction of Single Ligands and Pose Fitting of Native Ligands. For more accurate pose fitting and prediction, we tried POSIT v.1.0.2 from the OpenEye suite.[65] We prepared the receptors and allowed mild ligand−protein clashes in the generated receptors to account for the average coordinate error expected in PDB structures. We used the combine-receptor option to allow for identifying and subsequent use of pockets that would be unexplored if we used only single PDB files. The merged receptors were used in the FRED docking step as well. We allowed alternate posing of each ligand within 0.5 Å RMSD in each receptor. Mild clashes similar to those used in receptor preparation were allowed during pose prediction. We forced aromatic rings to be planar. The minimum probability to accept poses within 2.0 Å of the native ligand was set to 0.33 with minimum initial probability of 0.05. Receptors that had initial rigid TanimotoCombo < 0.8[66] were rejected. All generated protein−ligand complexes were subjected to a final optimization preserving the interactions

dx.doi.org/10.1021/ci400040d | J. Chem. Inf. Model. 2013, 53, 1934−1945

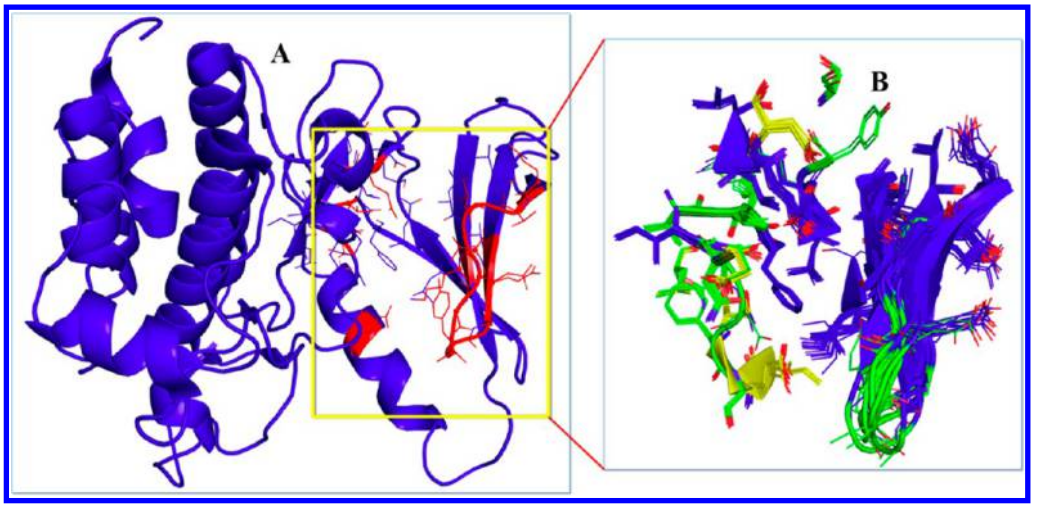


Figure 3. Structure of CHK1 (PDB ID: 2E9N, panel A) is displayed as blue cartoon. The most flexible residues have side chains represented as red lines (inside the yellow box). This flexibility is inferred from aligning the CHK1 active site from multiple PDB structures (panel B). The α-helices, β-sheets, and loops in panel B are yellow, blue, and green, respectively. Images were generated using PyMol.

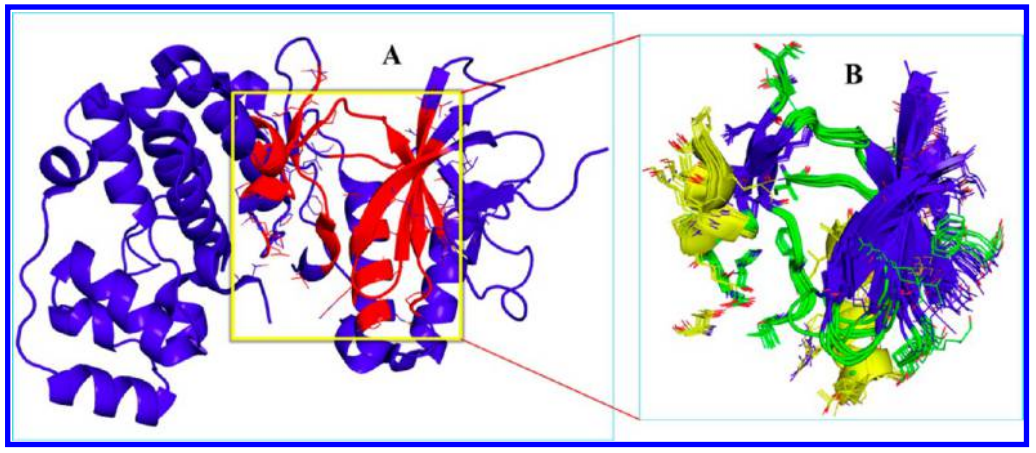


Figure 4. Structure of ERK2 (PDB ID: 3I5Z, panel A) is displayed as blue cartoon. The most flexible residues have side chains represented as red lines (inside the yellow box). The flexibility of the active site is shown by the alignment of the ERK2 binding site as determined from multiple PDB structures (panel B). The α-helices, β-sheets, and loops in panel B are yellow, blue, and green, respectively. Images were generated using PyMol.

associated with atoms involved in the TanimotoCombo score. A cutoff of 10 kcal/mol was used as the maximum strain to accept.

3. RESULTS AND DISCUSSION

Ligand recognition by a protein depends on both shape (3D structure) and electrostatic complementarities. Several reports have described the effect of ligand and protein preparation steps on molecular docking efficiency. We used LigPrep, Protein Preparation Wizard, and SPORES to adjust the protonation, tautomeric, and stereoisomeric states of protein and ligand databases. Ligand conformational sampling is as important as correct ligand preparation. FRED and HYBRID require multiconformer databases to be prepared separately. Glide modules—Glide docking, Glide ensemble docking, and Glide induced-fit docking (IFD)—PLANTS and POSIT each have an integrated conformational search algorithm. Pregeneration of conformers could be more beneficial if we have ligand databases with saturated rings, and non-specified E/Z vinyl, E/Z amide bonds, and stereogenic centers, and if the integrated conformational search algorithm has limited options to handle these situations. Molecular docking efficiency is influenced by the correct preparation of the target protein structure.[67] All PDB structures that were used in this study were checked for integrity, especially of the active site region, and any structure which showed gap(s) in the binding site region was rejected. Some reports advocate not to include any geometry refinement step prior to generating the receptor for docking because this may incorrectly improve the protein−ligand interactions.[68] Protein−ligand complexes were minimized before preparing the receptor grid for use in the Glide modules. This minimization step was not implemented for the other software applications. The active site was defined by the amino acid residues surrounding the bound ligand (Figure 2). Protein conformational changes often take place upon ligand binding, so ignoring protein flexibility during molecular docking may give results that are incorrect.[69] There are several approaches to include protein flexibility in the docking procedure. We tried rigid body docking with FRED, soft receptor docking by softening the potential of receptor nonpolar regions with Glide, docking using receptors with flexible side chains with PLANTS, docking in multiple protein structures with HYBRID and Glide ensemble docking, and IFD. Each technique has its own approximations, advantages, and limitations.

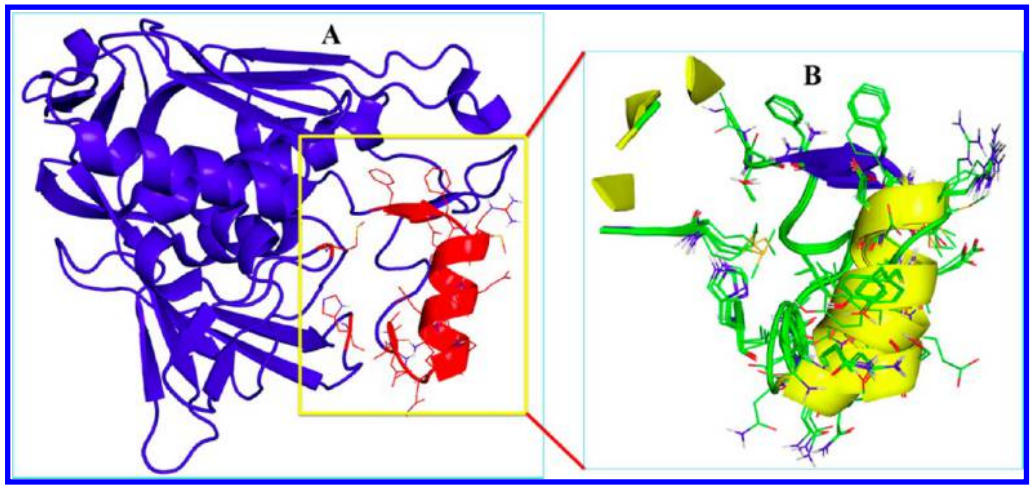


Figure 5. Structure of LpxC (PDB ID: 3P3E, panel A) is displayed as blue cartoon. The most flexible residues have side chains represented as red lines (inside the yellow box). The flexibility of the active site is shown by the alignment of the LpxC binding site as determined from multiple PDB structures (panel B). The α-helices, β-sheets, and loops in panel B are yellow, blue, and green, respectively. Images were generated using PyMol.

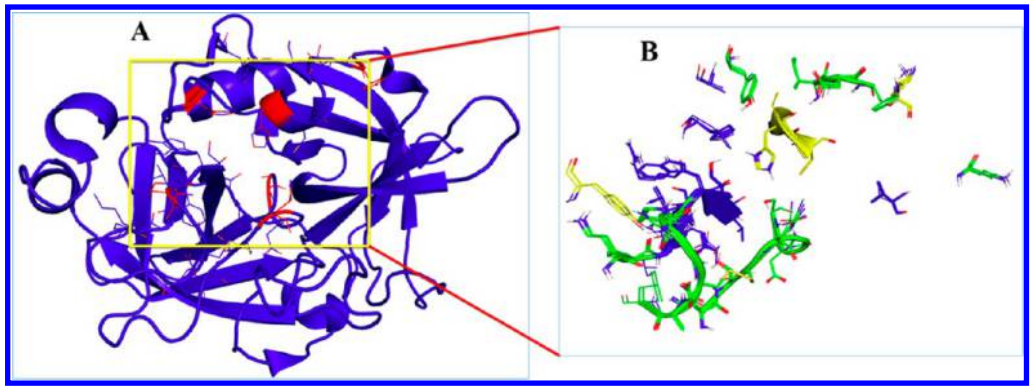


Figure 6. Structure of UPA (PDB ID: 1OWE, panel A) is displayed as blue cartoon. The most flexible residues have side chains represented as red lines (inside the yellow box). The flexibility of the active site is shown by the alignment of the UPA binding site as determined from multiple PDB structures (panel B). The α-helices, β-sheets, and loops in panel B are yellow, blue, and green, respectively. Images were generated using PyMol.

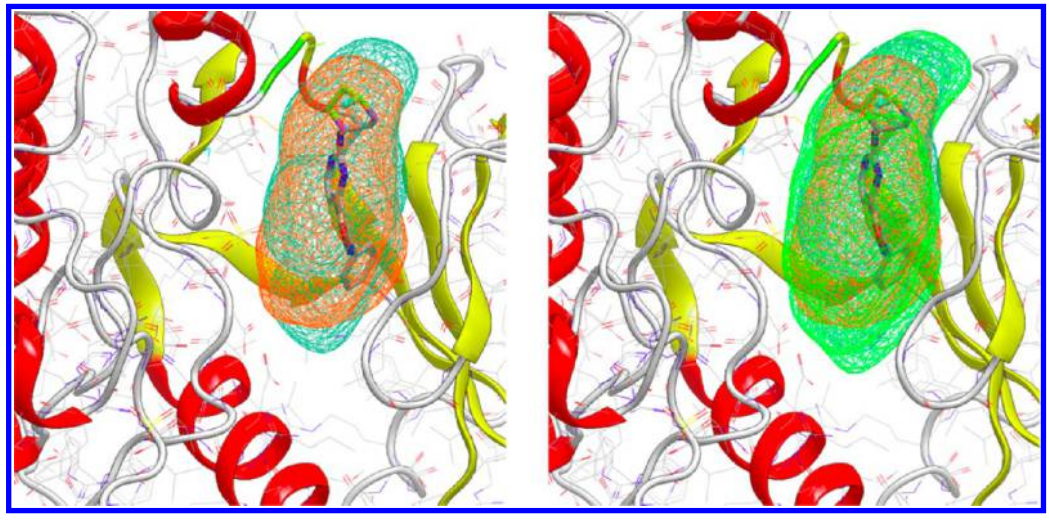


Figure 7. Binding area of ERK2 as defined by a single PDB structure, 3I5Z, is shown in orange surface mesh (around the ligand) and dark green surface mesh (extended to the protein surface close to the ligand) (left), and the merged area as defined by two PDB structures, 3I5Z and 4FUX, as shown in light green surface mesh surrounding the ligand and extended protein surface from 3I5Z alone (right). The α-helices, β-sheets, and loops are red, yellow, and white, respectively. Images were generated using VIDA.[73].

As a first step, we aligned the binding sites of the PDB structures of each target to check for conformational flexibility of active site residues. The four targets have different degrees of flexibility. CHK1 showed a high degree of flexibility in the P-loop region (residues 13−23) and for the side chains of Lys38, Glu55, Val68, Lys91, Ser147, and Asp148 (Figure 3). The P-

dx.doi.org/10.1021/ci400040d | J. Chem. Inf. Model. 2013, 53, 1934−1945

Table 1. Docking Algorithm Features, Computational Times, and Percent Accuracies

					ensemble docking	
	rigid	soft	FSC[a]	IFD	Glide	hybrid
scoring function	Chemgauss4	Glide SP	$PLANTS_{PLP}$	Glide SP	Glide SP	Chemgauss4
conformational sampling	pregeneration	integral	integral	integral	integral	pregeneration
comp. time/ligand (min)	<1	1−5	∼30	∼60	∼45	∼15
protein			% accuracy			
CHK1	11	4	4	4	10	21
ERK2	15	7	20	32	42	60
LpxC	81	79	1	73	75	84
UPA	1	10	51	43	1	25

[a]FSC is Flexible Side Chain.

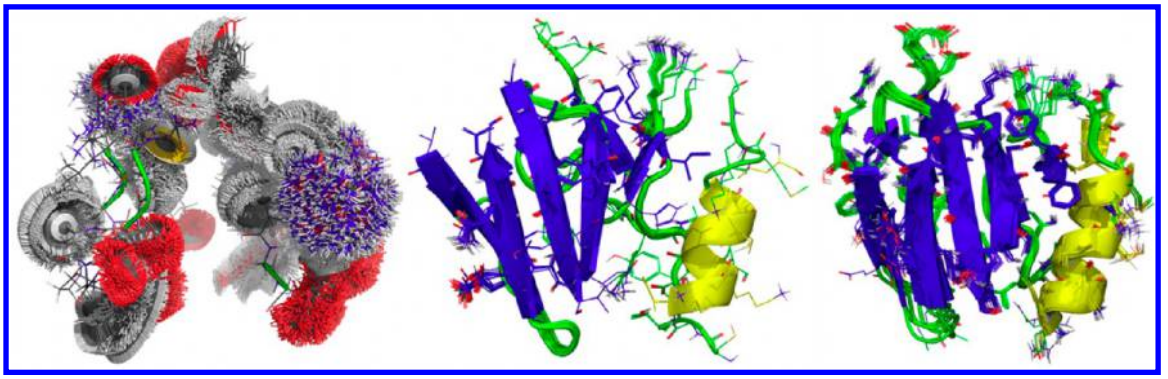


Figure 8. Binding region used in flexible docking to the CHK1 active site. (Left) PLANTS generated multiple orientations for each side chain of the active site residues. Each conformation was tried for each ligand, and the best scoring conformation was kept (carbon atoms are gray). (Middle) The IFD approach allowed for domain movements (backbones and side chain orientations) upon ligand binding, and the P-loop region was shown to be flexible to allow ligand fitting (carbon atoms are green). (Right) Ensemble docking used multiple PDB structures aligned based on the active site information. Each individual PDB structure was tried for each ligand, and the best scoring PDB structure was saved for each ligand separately (carbon atoms are green). In the center and right panels, the α-helices, β-sheets, and loops are yellow, blue, and green, respectively. Images were generated using PyMol.

loop in protein kinases often moves to accommodate ligands with variable sizes,[70] and the hydroxyl, carboxyl, and amino side chains of Lys, Glu, Val, and Asp in the active site rotate to allow for making hydrogen bonding and electrostatic interactions with bound ligands. In the case of ERK2, a high degree of flexibility is observed throughout the active site (Figure 4) that allows varied ligands to fit. LpxC showed some flexibility in residues 191−207, His19, and Met62 (Figure 5). The structure is relatively conformationally conserved at the metal binding site, as shown by the minor orientation differences for the histidine residue (Figure 5B). In the case of UPA, scattered residues are shown to be flexible around the active site (Figure 6). UPA active site flexibility mainly includes rotation of the side chains of the involved residues, such as His54, Tyr59, His106, Gln208, Ser211, and Arg233, to form hydrogen bonds with bound ligands.

Active site water molecules can be considered another important aspect of target flexibility. Water molecules should be checked carefully to avoid using artifact waters (those that are not essential to the protein structure) in the docking process. Using artifact active site water molecules can have a deleterious effect by providing false energetic stability to the protein−ligand complex. The PDB structures of CHK1, ERK2, LpxC, and UPA, which were used in the initial docking step, were checked carefully for active site water molecules. CHK1, ERK2, LpxC, and UPA have, respectively, 2, 2, 6, and 7 active site water molecules showing at least two hydrogen bonds with non-waters (with protein and/or ligand).

The binding site area that is defined by one PDB structure does not necessarily have the same features as the binding site area of another PDB structure of the same protein. To avoid misleading information that can come from relying on a single PDB structure, which would be expected to degrade the docking performance, we compared several protein−ligand co-crystal structures of each target. We explored the binding area of each target (Figure 7) to make use of all possible pockets in the rigid docking step. We tried receptor merging to add the advantage of fast rigid docking to the use of an expanded protein conformational space.

The performance variability of docking software applications (Table 1) may be attributed to the specific target at hand, the scoring function, and the ligand and protein conformational sampling approaches. We defined percent accuracy for a particular method as 100 times the correlation coefficient, r^2, between docking scores and experimental activity data (Supporting Information). The CHK1 active site has a high degree of flexibility, with the P-loop residues showing the greatest conformational changes upon ligand binding (induced fit). We expected the flexible docking approaches (Figure 8) to perform better in this case; however, the rigid-body docking (Figure 9) showed the best accuracy. Allowing adjustments to the receptor conformation through the soft receptor approach, using flexible side chains or IFD, did not show any improvement. Ligand sampling is a key factor in this case. OEDocking used OMEGA ligand conformational sampling as a predocking step that allowed adding more flexibility during

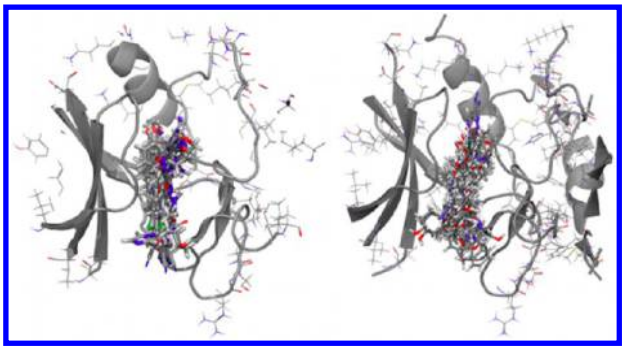


Figure 9. Docked poses in the CHK1 active site using rigid receptor approaches. (Left) OEDocking with pre-generation of ligand conformers using OMEGA. (Right) Soft receptor approach using Glide Maestro with self-generation of ligand conformers. Images were generated in Maestro.[74] All secondary structures are colored gray; ligand poses are displayed as sticks (with carbon gray, oxygen red, and nitrogen blue).

conformer generation. In other software, ligand sampling is an internalized feature with limited ability to be manipulated. Protein sampling in the flexible docking approaches could not provide the required flexibility to handle the highly movable active site regions. Also, the conformational space occupied by the binding pocket upon binding with some of the varied ligands is not covered by the available crystal structures. Using multiple protein structures in Glide and OEDocking gave a moderate increase in the docking performance from 4% to 10% and from 11% to 21%, respectively. The larger improvement in the case of OEDocking is likely because of the predocking ligand sampling and the approach of using an ensemble of structures. In cases like that of CHK1, docking techniques cannot be depended on to play a significant role in the virtual screening process.

The active site of ERK2 contains residues that are involved in forming parts of the protein's β-sheets, α-helices, and loops (Figure 4). PLANTS and IFD both showed that the active site is highly flexible (Figure 10). The observed high flexibility of backbone and side chains of the active site residues would be expected to limit the docking efficiency. For this protein target, there were substantial increases in the accuracy when using the flexible techniques. Combining the predocking ligand sampling with the use of an ensemble of structures showed the best accuracy (60%). The conformational space provided by the multiple crystal structures was varied enough to cover the conformational space of the ERK2 active site and, hence, to improve the ensemble docking performance.

The active site of LpxC contains Zn^{2+}, and there is structural evidence that ligands form metal chelates upon binding. Adding a chelate formation as a docking constraint enhanced the overall accuracy of all techniques (Figure 11) except for the flexible side chain method, which needs special handling of this metalloenzyme before running docking. PLANTS generated a large number of side chain conformers throughout the active site even though there is supposed to be a more conserved region at the metal binding site (Figure 5), explaining why PLANTS performed badly for this target. The IFD approach allowed for domain movements but did not move the metal-binding side chains.

For UPA, the active site showed little flexibility according to the evidence from the available crystal structures. Hence, we anticipated rigid docking methods to perform well; however, they showed the lowest accuracy. The flexible side chain method and IFD (Figure 12) showed the best performance, providing enough space for ligands to fit. Upon combining predocking conformational sampling with using multiple crystal structures, the results showed a significant increase in performance from 1% to 25%. The conformational flexibility of the active site of UPA is not covered by the available crystal structures, and this is the main reason for the low accuracy of the ensemble docking methods.

The overall performance of rigid body docking was increased by using merged receptors over using the single receptor, but that still was not as good as using multiple receptors. The merging method adds more information to the receptor model and provides additional pockets that were unexplored by the original receptor. As we previously described, active site waters should be manipulated with extreme care. Ligands may interact with real active site water molecules forming hydrogen bonds or may displace active site waters, disrupting an important hydrogen bond network. Artifact water molecules caused by the crystallization technique will falsely appear to improve the energetics of ligand binding and the docking score. The latter

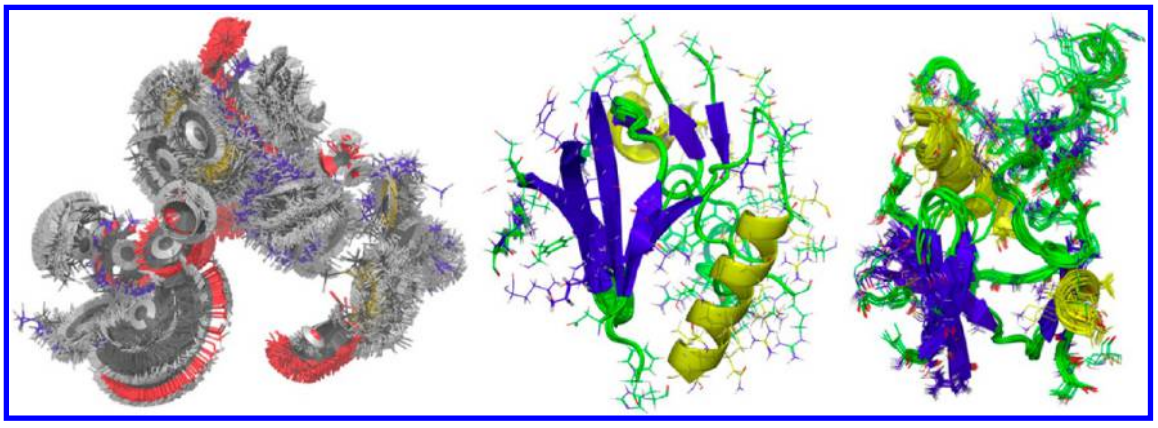


Figure 10. Docking site (active site) of ERK2. (Left) PLANTS generated multiple orientations for each side chain of the active site residues (carbon atoms are gray). (Middle) The IFD approach allowed for backbones and side chain orientations to be adjusted to accommodate ligand binding (carbon atoms are green). (Right) Multiple PDB structures were used for the ensemble docking and aligned based on the active site information (carbon atoms are green). In the center and right panels, the α-helices, β-sheets, and loops are yellow, blue, and green, respectively. Images were generated using PyMol.

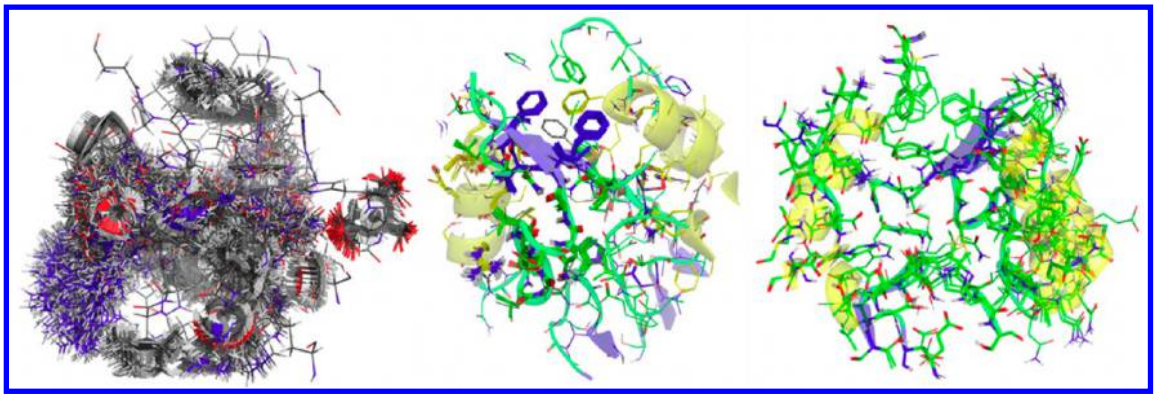


Figure 11. Docking site (active site) of LpxC. (Left) PLANTS generated multiple orientations for each side chain of the active site residues (carbon atoms are gray). (Middle) The IFD approach allowed for adjustments to accommodate ligand binding (carbon atoms are green). (Right) Ensemble docking used multiple PDB structures and aligned them based on the active site information (carbon atoms are green). In the center and right panels, the α-helices, β-sheets, and loops are yellow, blue, and green, respectively. Images were generated using PyMol.

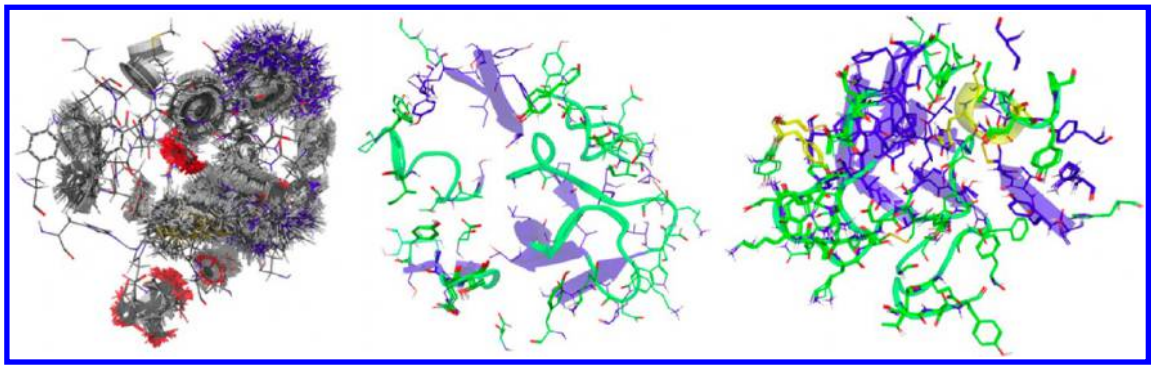


Figure 12. Docking site (active site) of UPA. (Left) PLANTS generated multiple orientations for each side chain of the active site residues (carbon atoms are gray). (Middle) The IFD approach allowed for adjustments to accommodate ligand binding (carbon atoms are green). (Right) Ensemble docking used multiple PDB structures and aligned them based on the active site information. In the center and right panels, the α-helices, β-sheets, and loops are yellow, blue, and green, respectively. Images were generated using PyMol.

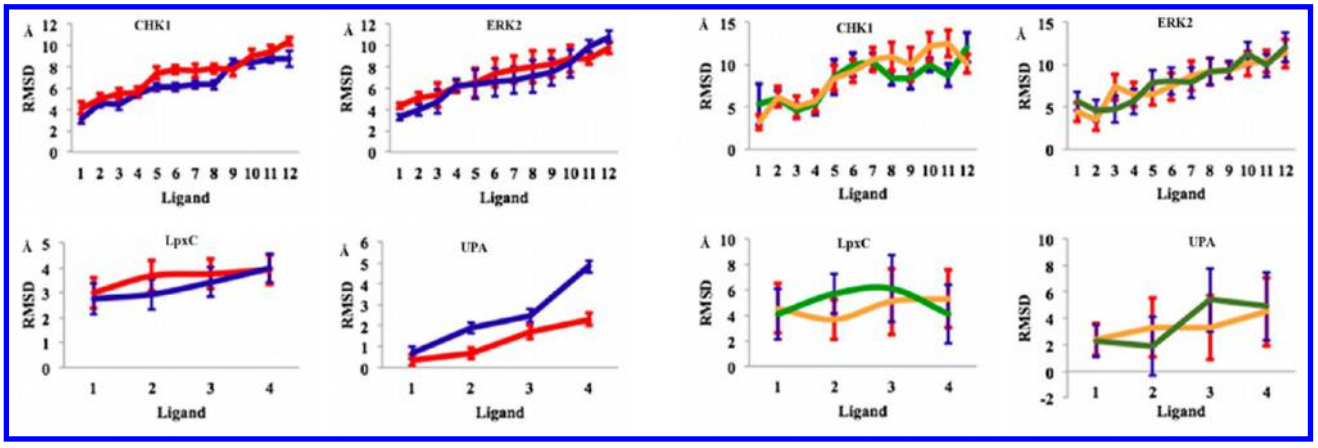


Figure 13. Performance of docking applications to regenerate correct ligand poses. Ligands (ordered from left to right based on increasing RMSD) are plotted against their ligand RMSD values. Each line represents the ligand RMSD, while the "error bars" represent the magnitude of the protein RMSD (cf. main text). Blue, rigid docking (OEDocking's FRED); red, soft receptor docking (Schrödinger's Glide); green, IFD (Schrödinger); orange, flexible side chain docking (PLANTS).

scenario is shown to occur in this case. The number of surviving active site water molecules during protein preparation was limited by keeping only waters forming two hydrogen bonds with non-waters. When we increased the number of constraints such that waters needed to form three or four hydrogen bonds with non-waters, none of the proteins had any water molecules in the active site except for LpxC, which showed one water molecule in the active site. Incorporation of active site waters led to a significant decrease in all docking performance.

We checked the performance of the docking applications to regenerate the correct pose of ligands of already solved X-ray co-crystal structures by performing docking using any ligands for which there were such structures, from the training or test

dx.doi.org/10.1021/ci400040d | J. Chem. Inf. Model. 2013, 53, 1934−1945

sets. The docking was not done to the ligand's native X-ray crystal structure ("redocking") but rather to a different X-ray structure of the same protein. This docking was done as part of the primary study before the experimental activity data in the CSAR exercise was released. Then comparison was made between the final docked pose and the co-crystallized docked conformation of the same ligand. In order to do this it was necessary to align the two protein structures first (all protein atoms were used in this rigid alignment). To check the performance of regenerating the correct ligand pose, we docked ligands having known protein bound crystal structures into 2E9N (CHK1), 3I5Z (ERK2), 3P3E (LpxC), and 1OWE (UPA). We calculated the RMSD (for all protein atoms only) for each of the PDB structures used in the docking step compared to the original PDB co-crystallized structure of the particular ligand. We referred to this value as the protein RMSD for a particular ligand. Next, the difference for the ligand only between the docking pose and the crystal conformation was measured by calculating an atom-by-atom ligand RMSD. The rigid docking approach we used for this (OEDocking's FRED) out-performed the soft receptor approach (Schrödinger's Glide) in three of the four cases, other than for UPA (Figure 13). In flexible receptor methods, the side chains and protein backbone can move and/or rotate to help a particular ligand to fit more optimally into the binding pocket, but the movement of the protein may result in a deviation of the ligand from the actual binding pose found in the crystal structure.[22] To make sure we were using a common frame for comparing ligands' RMSD in case of flexible techniques, we used the RMSD of protein backbone atoms instead of all atoms as in the previous case (Figure 13). The higher values of protein RMSD in flexible approaches is attributed to the allowed movements of side chain and backbone active site residues that lead to considerable deviations form the original crystal structure. Because of residue movements, the best scoring docked pose in most cases is not the one with the lowest ligand RMSD value. Rigid body docking and soft receptor methods showed better performance than the flexible receptor methods, especially if these approaches were combined with using multiple receptors in the hybrid and ensemble docking algorithms.

We used POSIT for docking of all ligands and compared those for which there is an X-ray co-crystal structure to those for which there is none. The former had docked poses with RMSD < 1.0 Å compared to their crystal structures. Posit predicted for the active ligands with no X-ray co-crystal structure that they had probabilities >70% to bind within the active site of the proteins. Greater than 90% of the inactive ligands did not show any binding probability.

The scoring functions are another important aspect to address because they play an important role in ligand ranking and pose selection. In OEDocking, the scoring function is used to select the best pose and ligand placement in the active site is based on a shape-fitting algorithm. In Glide and flexible side chain algorithms, ligand posing and ranking are based solely on the scoring function. We found that OEDocking performed the best. Utilization of the newly implemented chemical Gaussian overlay (CGO) function[71] may enhance OEDocking performance even more.

4. CONCLUSION

The best docking technique should be chosen after studying in detail the target, candidate ligands, and docking method performance. Benchmark analysis should be considered before choosing the technique. Protein flexibility could be considered based on the facts of the case under consideration. The rigid receptor method showed high accuracy for ranking active ligands. It performed better in case of metalloenzymes than in the other cases. Soft receptor methods were comparable to rigid body docking, with better performance in UPA. The flexible side chain method had moderate performance in most cases and performed better in UPA. It gave the worst results in the case of the metalloenzyme. It needs special handling of the target before running docking. The induced fit docking (IFD) method showed stable results in all cases. By modifying the softening potential, IFD produced enough flexibility to adjust the p-loop of ERK2, providing a better representation for the docking step and hence a better chance for improved results. Ensemble docking methods showed stable results as well, except for UPA. The number of crystal structures of UPA was not enough to cover the conformational space of the enzyme. Computational expense is an important issue, especially for virtual screening of a large number of candidate ligands. IFD is the most expensive of the techniques, in particular if we use the Glide XP scoring function. Glide ensemble docking would be the most expensive if we included molecular dynamics and MM-GBSA calculations for obtaining more accurate and representative binding free energies. Incorporation of structural water should be considered only after careful analysis. In this work, incorporation of active site waters negatively affected the results of all docking methods and in particular lowered the performance of the IFD and flexible side chain methods.

In general, we may summarize our major findings with the following points:

1. It is better to separately generate the ligand conformers as efficiently as possible and not depend upon the self-generation approach.
2. For the protein preparation step, it is preferable not to minimize the complex because this will bias the protein−ligand interaction profile and hence will affect the docking results.
3. Careful analysis of active site crystal water molecules is required. Inclusion of water molecules should be considered after studying the hydrogen bonding with non-water residues and after studying the relative abundance of water molecules by analysis of multiple crystal structures.
4. If the target under consideration has multiple crystal structures with a good coverage of possible active site conformations, the hybrid approach with rigid receptors or the ensemble docking protocol with soft receptors will be preferred due to their accuracy and computational efficiency.
5. If the target does not have multiple crystal structures and there is prior knowledge from benchmark studies of possible movements of the active site residues, IFD and FSC protocols should be considered.
6. For virtual screening purposes, it is better to consider pose fitting and prediction approaches to rule out structures that do not bind in the same manner as native ligands. Only those compounds that bind in a similar way to that of the native ligands would be used in further steps of virtual screening.

■ ASSOCIATED CONTENT

Supporting Information

Docking scores (*y*-axis) versus reported experimental activity data for each of the four targets and six docking algorithms. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rjd@olemiss.edu. Phone: 662-915-5880. Fax: 662-915-5638. Address: 21 Faser Hall, Department of Medicinal Chemistry, University of Mississippi, University, Mississippi, 38677-1848.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Thanks to Ronak Y. Patel for helpful discussions. This work was supported in part by the National Science Foundation under Awards EPS 0903787 and EPS 1006883. This investigation was conducted in part in a facility constructed with support from research facilities improvement program C06 RR-14503-01 from the NIH NCRR.

■ REFERENCES

(1) Huang, S.-Y.; Zou, X. Advances and challenges in protein−ligand docking. *Int. J. Mol. Sci.* **2010**, *11* (8), 3016−3034.

(2) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein−ligand docking: Current status and future challenges. *Proteins: Struct., Funct., Bioinf.* **2006**, *65* (1), 15−26.

(3) Novikov, F. N.; Chilov, G. G. Molecular docking: Theoretical background, practical applications and perspectives. *Mendeleev Commun.* **2009**, *19* (5), 237−242.

(4) Balaji, G. A.; Balaji, V. N.; Rao, S. N. Utility of scoring function customization in docking-based virtual screening approaches. *Curr. Sci.* **2013**, *104* (1), 86−97.

(5) 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.

(6) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. *Arch. Biochem. Biophys.* **1978**, *185* (2), 584−591.

(7) RCSB Protein Data Bank (PDB). http://www.rcsb.org/ (accessed November 16, 2012).

(8) Mario Geysen, H.; Schoenen, F.; Wagner, D.; Wagner, R. Combinatorial compound libraries for drug discovery: An ongoing challenge. *Nat. Rev. Drug Discovery* **2003**, *2* (3), 222−230.

(9) Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J. L.; Martínez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D. K. Recognizing pitfalls in virtual screening: A critical review. *J. Chem. Inf. Model.* **2012**, *52* (4), 867−881.

(10) Onodera, K.; Satou, K.; Hirota, H. Evaluations of molecular docking programs for virtual screening. *J. Chem. Inf. Model.* **2007**, *47* (4), 1609−1618.

(11) Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. Comparison of several molecular docking programs: Pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2009**, *49* (6), 1455−1474.

(12) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian docking functions. *Biopolymers* **2003**, *68* (1), 76−90.

(13) Jain, A. N. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, *46* (4), 499−511.

(14) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739−1749.

(15) Thomsen, R.; Christensen, M. H. MolDock: A new technique for high-accuracy molecular docking. *J. Med. Chem.* **2006**, *49* (11), 3315−3321.

(16) McGann, M. FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2011**, *51* (3), 578−596.

(17) Neves, M.; Totrov, M.; Abagyan, R. Docking and scoring with ICM: The benchmarking results and strategies for improvement. *J. Comput.-Aided Mol. Des.* **2012**, *26* (6), 675−686.

(18) McGaughey, G. B.; Sheridan, R. P.; Bayly, C. I.; Culberson, J. C.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J.-F.; Cornell, W. D. Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* **2007**, *47* (4), 1504−1519.

(19) Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245* (1), 43−53.

(20) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267* (3), 727−748.

(21) Korb, O.; Cole, J., Ant Colony Optimization for Ligand Docking. In *Proceedings of the 7th International Conference on Swarm Intelligence*; Springer-Verlag: Brussels, Belgium, 2010; pp 72−83.

(22) Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337* (1), 209−225.

(23) Barreca, M. L.; Iraci, N.; De Luca, L.; Chimirri, A. Induced-fit docking approach provides insight into the binding mode and mechanism of action of HIV-1 integrase inhibitors. *ChemMedChem* **2009**, *4* (9), 1446−1456.

(24) Davis, I. W.; Baker, D. Rosetta ligand docking with full ligand and receptor flexibility. *J. Mol. Biol.* **2009**, *385* (2), 381−392.

(25) Huang, S.-Y.; Zou, X. Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins: Struct., Funct., Bioinf.* **2007**, *66* (2), 399−421.

(26) Huang, S.-Y.; Grinter, S. Z.; Zou, X. Scoring functions and their evaluation methods for protein−ligand docking: recent advances and future directions. *Phys. Chem. Chem. Phys.* **2010**, *12* (40), 12899−12908.

(27) Yin, S.; Biedermannova, L.; Vondrasek, J.; Dokholyan, N. V. MedusaScore: An accurate force field-based scoring function for virtual drug screening. *J. Chem. Inf. Model.* **2008**, *48* (8), 1656−1662.

(28) Huang, N.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. Physics-based scoring of protein−ligand complexes: Enrichment of known inhibitors in large-scale virtual screening. *J. Chem. Inf. Model.* **2005**, *46* (1), 243−253.

(29) Böhm, H.-J. The development of a simple empirical scoring function to estimate the binding constant for a protein−ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8* (3), 243−256.

(30) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11* (5), 425−445.

(31) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein−ligand interactions: A simplified potential approach. *J. Med. Chem.* **1999**, *42* (5), 791−804.

(32) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein−ligand interactions. *J. Mol. Biol.* **2000**, *295* (2), 337−356.

(33) Ruvinsky, A. Calculations of protein-ligand binding entropy of relative and overall molecular motions. *J. Comput.-Aided Mol. Des.* **2007**, *21* (7), 361−370.

(34) Chang, M. W.; Belew, R. K.; Carroll, K. S.; Olson, A. J.; Goodsell, D. S. Empirical entropic contributions in computational

docking: Evaluation in APS reductase complexes. *J. Comput. Chem.* **2008**, *29* (11), 1753−1761.
(35) Lee, J.; Seok, C. A statistical rescoring scheme for protein−ligand docking: Consideration of entropic effect. *Proteins: Struct., Funct., Bioinf.* **2008**, *70* (3), 1074−1083.
(36) Liu, S.; Fu, R.; Zhou, L.-H.; Chen, S.-P. Application of consensus scoring and principal component analysis for virtual screening against β-secretase (BACE-1). *PLoS One* **2012**, *7* (6), e38086.
(37) Feher, M. Consensus scoring for protein−ligand interactions. *Drug Discovery Today* **2006**, *11* (9−10), 421−428.
(38) Teramoto, R.; Fukunishi, H. Consensus scoring with feature selection for structure-based virtual screening. *J. Chem. Inf. Model.* **2008**, *48* (2), 288−295.
(39) Ross, G. A.; Morris, G. M.; Biggin, P. C. Rapid and accurate prediction and scoring of water molecules in protein binding sites. *PLoS One* **2012**, *7* (3), e32036.
(40) 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* (4), 997−1009.
(41) Tong, Y.; Claiborne, A.; Stewart, K. D.; Park, C.; Kovar, P.; Chen, Z.; Credo, R. B.; Gu, W.-Z.; Gwaltney, S. L., II; Judge, R. A.; Zhang, H.; Rosenberg, S. H.; Sham, H. L.; Sowin, T. J.; Lin, N.-H. Discovery of 1,4-dihydroindeno[1,2-c]pyrazoles as a novel class of potent and selective checkpoint kinase 1 inhibitors. *Bioorg. Med. Chem.* **2007**, *15* (7), 2759−2767.
(42) Community Structure−activity Resource (CSAR). http://www.csardock.org/ (accessed April 2, 2013).
(43) Aronov, A. M.; Tang, Q.; Martinez-Botella, G.; Bemis, G. W.; Cao, J.; Chen, G.; Ewing, N. P.; Ford, P. J.; Germann, U. A.; Green, J.; Hale, M. R.; Jacobs, M.; Janetka, J. W.; Maltais, F.; Markland, W.; Namchuk, M. N.; Nanthakumar, S.; Poondru, S.; Straub, J.; ter Haar, E.; Xie, X. Structure-guided design of potent and selective pyrimidylpyrrole inhibitors of extracellular signal-regulated kinase (ERK) using conformational control. *J. Med. Chem.* **2009**, *52* (20), 6362−6368.
(44) Lee, C.-J.; Liang, X.; Chen, X.; Zeng, D.; Joo, S. H.; Chung, H. S.; Barb, A. W.; Swanson, S. M.; Nicholas, R. A.; Li, Y.; Toone, E. J.; Raetz, C. R. H.; Zhou, P. Species-specific and inhibitor-dependent conformations of LpxC: Implications for antibiotic design. *Chem. Biol.* **2011**, *18* (1), 38−47.
(45) Wendt, M. D.; Rockway, T. W.; Geyer, A.; McClellan, W.; Weitzberg, M.; Zhao, X.; Mantei, R.; Nienaber, V. L.; Stewart, K.; Klinghofer, V.; Giranda, V. L. Identification of novel binding interactions in the development of potent, selective 2-naphthamidine inhibitors of urokinase. Synthesis, structural analysis, and SAR of N-phenyl amide 6-substitution. *J. Med. Chem.* **2003**, *47* (2), 303−324.
(46) *Schrödinger Suite 2012: Protein Preparation Wizard*; Schrödinger, LLC: New York.
(47) *Schrödinger Suite 2012*; Schrödinger, LLC: New York.
(48) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. A hierarchical approach to all-atom protein loop prediction. *Proteins: Struct., Funct., Bioinf.* **2004**, *55* (2), 351−367.
(49) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.* **2002**, *320* (3), 597−608.
(50) *Schrödinger Suite 2012: Prime*, version 3.1; Schrödinger, LLC: New York.
(51) Rostkowski, M.; Olsson, M.; Sondergaard, C.; Jensen, J. Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. *BMC Structural Biology* **2011**, *11* (1), 6.
(52) Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (19), 6665−6670.
(53) *Schrödinger Suite: LigPrep*, version 2.5; Schrödinger, LLC: New York, 2012.
(54) *OEDocking v3.0.0: FRED v3.0.0*; OpenEye Scientific Software: Santa Fe, NM. http://www.eyesopen.com (accessed April 2, 2013).
(55) *OEDocking v3.0.0: HYBRID v3.0.0*; OpenEye Scientific Software: Santa Fe, NM. http://www.eyesopen.com (accessed April 2, 2013).
(56) *OMEGA v2.4.6*; OpenEye Scientific Software: Santa Fe, NM. http://www.eyesopen.com (accessed April 2, 2013).
(57) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. *J. Chem. Inf. Model.* **2010**, *50* (4), 572−584.
(58) Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20* (7), 720−729.
(59) *Schrödinger Suite 2012: Glide*, version 5.8; Schrödinger, LLC: New York, 2012.
(60) Korb, O.; Stützle, T.; Exner, T. E. *Protein−Ligand ANT System*, version 1.2; Universität Konstanz, Fachbereich Chemie und Zukunftskolleg: Konstanz, Germany, 2012.
(61) Zonta, N. *ZODIAC*, version 0.65 beta. http://www.zeden.org (accessed April 2, 2013).
(62) ten Brink, T.; Exner, T. E. *Structure PrOtonation and REcognition System (SPORES)*, version 1.28; Universität Konstanz, Fachbereich Chemie und Zukunftskolleg: Konstanz, Germany, 2012.
(63) Korb, O.; Stüzle, T.; Exner, T., PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design. In *Ant Colony Optimization and Swarm Intelligence*; Dorigo, M., Gambardella, L., Birattari, M., Martinoli, A., Poli, R., Stützle, T., Eds.; Springer: Berlin, Heidelberg, 2006; Vol. *4150*, pp 247−258.
(64) Korb, O.; Stützle, T.; Exner, T. E. Empirical scoring functions for advanced protein−ligand docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49* (1), 84−96.
(65) *POSIT*, version 1.0.2; OpenEye Scientific Software: Santa Fe, NM. http://www.eyesopen.com accessed April 2, 2013).
(66) Tuccinardi, T.; Botta, M.; Giordano, A.; Martinelli, A. Protein kinases: Docking and homology modeling reliability. *J. Chem. Inf. Model.* **2010**, *50* (8), 1432−1441.
(67) ten Brink, T.; Exner, T. E. Influence of protonation, tautomeric, and stereoisomeric states on protein−ligand docking results. *J. Chem. Inf. Model.* **2009**, *49* (6), 1535−1546.
(68) Waszkowycz, B.; Clark, D. E.; Gancia, E. Outstanding challenges in protein−ligand docking and structure-based virtual screening. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1* (2), 229−259.
(69) Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: A practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18* (2), 178−184.
(70) Patel, R. Y.; Doerksen, R. J. Protein kinase-inhibitor database: Structural variability of and inhibitor interactions with the protein kinase P-loop. *J. Proteome Res.* **2010**, *9* (9), 4433−4442.
(71) McGann, M. FRED and HYBRID docking performance on standardized datasets. *J. Comput.-Aided Mol. Des.* **2012**, *26* (8), 897−906.
(72) *The PyMOL Molecular Graphics System*, version 1.3; Schrödinger, LLC: New York.
(73) *VIDA*, version 4.2.1; OpenEye Scientific Software: Santa Fe, NM. http://www.eyesopen.com accessed April 2, 2013).
(74) *Schrödinger Suite 2012: Maestro*, version 9.3.5; Schrödinger, LLC: New York.