JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

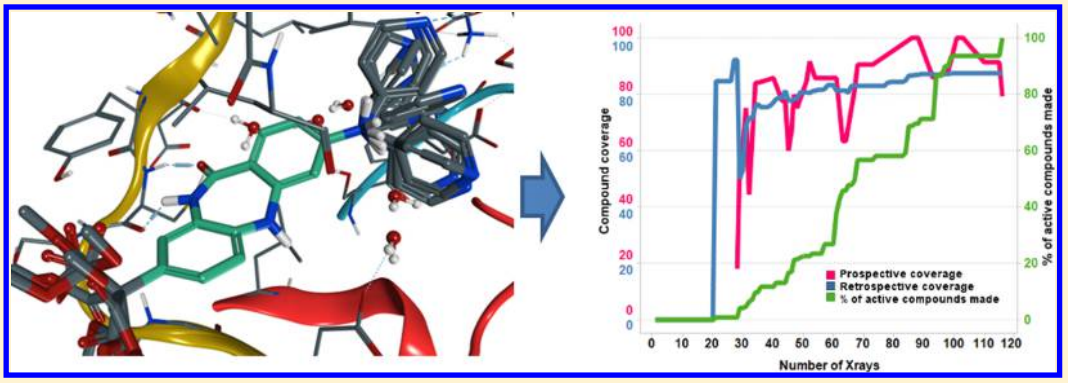
Knowledge-Based Strategy to Improve Ligand Pose Prediction Accuracy for Lead Optimization

Cen Gao,*[,†] Nels Thorsteinson,[§] Ian Watson,[‡] Jibo Wang,[†] and Michal Vieth[†]

[†]Discovery Chemistry, [‡]Advanced Analytics, Lilly Research Laboratories, Lilly Corporate Center, Indianapolis, Indiana 46285, United States

[§]Chemical Computing Group Inc., 1010 Sherbrooke St. W, Suite 910, Montreal, Quebec H3A 2R7, Canada

Supporting Information



ABSTRACT: Accurately predicting how a small molecule binds to its target protein is an essential requirement for structure-based drug design (SBDD) efforts. In structurally enabled medicinal chemistry programs, binding pose prediction is often applied to ligands after a related compound's crystal structure bound to the target protein has been solved. In this article, we present an automated pose prediction protocol that makes extensive use of existing X-ray ligand information. It uses spatial restraints during docking based on maximum common substructure (MCS) overlap between candidate molecule and existing X-ray coordinates of the related compound. For a validation data set of 8784 docking runs, our protocol's pose prediction accuracy (80−82%) is almost two times higher than that of one unbiased docking method software (43%). To demonstrate the utility of this protocol in a project setting, we show its application in a chronological manner for a number of internal drug discovery efforts. The accuracy and applicability of this algorithm (>70% of cases) to medicinal chemistry efforts make this the approach of choice for pose prediction in lead optimization programs.

1. INTRODUCTION

Identification and optimization of lead molecules are key steps in target-based drug design. In these stages, one or several chemical series (i.e., scaffolds) are optimized for their affinity, selectivity, pharmacokinetic profile, and other non-efficacy related properties. The optimization process is an iterative process of design, synthesis, and biological evaluation of a series of compounds. With the advent of X-ray crystallography, the acquisition of protein structures with bound lead molecules has become increasingly routine in many discovery efforts.[1] In addition to crystallized ligands, computationally predicted binding poses for structurally related ligands can greatly help interpreting structure−activity relationships (SARs) and rationalize selectivity trends.[2,3] More importantly, computational predictions of binding modes can be applied to evaluate the compounds designed by medicinal chemists. Multiple docking algorithms (e.g., GOLD,[4,5] GLIDE,[6] DOCK[7]) have been used in industrial and academic settings to predict binding modes and, in some cases, to virtually screen for new ligands. The pose prediction accuracy, i.e. comparison of predicted with experimentally observed poses, was found to be highly dependent on the protein target, the complexity of the ligands, as well as the sampling algorithms and scoring functions.[8−10] A recent literature survey[11] estimated that average cross-docking performance (measured by <2 Å RMSD between docked and X-ray coordinates) varied from 26% to 63% for data sets of significant size.

If a docking program cannot reproduce the existing binding pose, the docking model is often augmented by the addition of experimentally derived information. Various spatial constraints, such as substructure and shape, can greatly reduce the search space and bias sampling toward the correct solution. In many cases, these constrained models can improve the docking accuracy for a particular series. Automating the use of such constraints can be difficult, since manual intervention is often required for each new chemical series. Furthermore, in certain docking programs, constraints are implemented as filters to

Received: April 3, 2015
Published: June 19, 2015

DOI: 10.1021/acs.jcim.5b00186
J. Chem. Inf. Model. 2015, 55, 1460−1468

eliminate bad solutions rather than to reduce the search space toward good solutions. With the use of filter-based constraints, docking remains strongly dependent on the conformer search method utilized by each program and can often result in no solutions satisfying the imposed filter criteria. This becomes a notable issue for saturated- and macrocyclic-ring containing molecules due to the inherent difficulty in sampling these systems appropriately.

On the other hand, if constraints are applied to restrict the search to conformations observed in previous crystal structures (including unique ring arrangements) rather than filtering conformations generated in a non-biased manner during the docking protocol, accuracy will be improved. In fact, this is common practice for most structure-enabled medicinal chemistry efforts, where in contrast to virtual screening, computational involvement aids the chemistry effort in optimization of a few congeneric series (i.e., scaffolds) where X-ray structures exist to provide foundational understanding on binding of related ligands. By leveraging existing crystal structure and SAR knowledge, we often find "manual docking" can predict ligand binding poses in a very accurate manner. In the SAMPL1 blind docking test, several manual docking approaches performed particularly well when compared to sophisticated and fully automated docking programs.[12] Automation of this process can provide the benefits of accuracy while reducing expert time involved in manual steps. One idea introduced by Hare et al. in 2004, and utilized in the CORES algorithm is to use molecular frameworks of known ligands as a template to bias the docking process.[13] It was estimated that ~70% of the kinase ligands published between 1993 and 2002 in *The Journal of Medicinal Chemistry* can be modeled using related structures from the Protein Data Bank (PDB)[14] prior to the publication. Around the same time, we introduced SDOCKER, which combined an energy function with ligand shape data extracted from existing co-crystal structures.[15] The program automatically applied a shape penalty to ligands that move outside the region where the co-crystal ligands bind. This approach demonstrated overall improvement in docking accuracy compared to traditional docking methods.[16] The strategy of using binding modes derived from previously solved protein−ligand X-ray structures was further demonstrated in the de novo design of several potent and novel kinase libraries.[17] This idea is particularly useful in transferring binding mode information in target families, such as kinases. In fact, conservation of binding modes in kinases can be very helpful in fragment-based design methods.[18] In POSIT,[19] a docking program recently developed by OpenEye, an initial pose is generated based on shape or maximum common substructure (MCS) overlays to existing structures and followed by shape optimization or rigid docking. The final POSIT score is a success probability estimated using only the template ligand conformation. The method showed a superior result to other docking programs for a kinase set.[20]

Inspired by these approaches, we recently implemented an automated template-guided docking protocol. The workflow makes an extensive use of existing ligand-protein co-crystal structures by utilizing positional restraints in the MCS region in all docking stages. In this paper, we discuss the detailed implementation and validation of the workflow. We also present analysis, insights and classification of failed cases. In addition, we use the data from a set of Lilly's in-house medicinal chemistry programs to investigate the applicability of our approach in the course of drug discovery projects. The approach presented in this paper can, in principle, be used with any docking algorithm that allows the use of positional biasing restraints.

2. MATERIALS AND METHODS

2.1. MCS Docking Protocol. The workflow (Figure 1) can be divided into four steps: identification of candidate protein−ligand complexes, biased ligand conformational sampling, pose placement and clustering, and in-pocket refinement and final pose ranking.

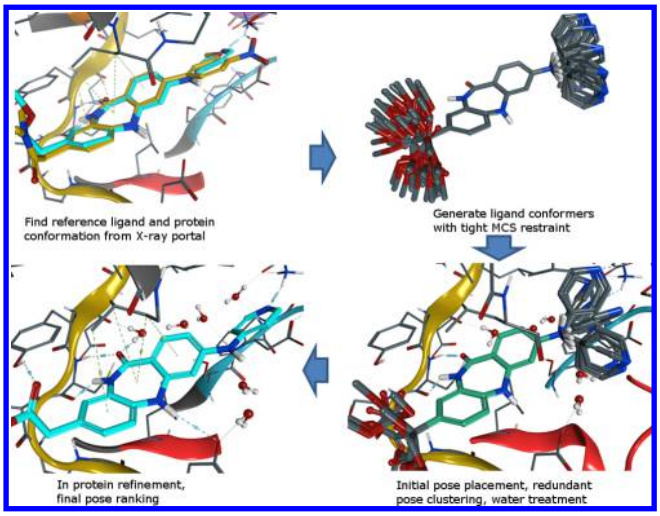


Figure 1. MCS docking protocol.

For each candidate ligand, a chirality-sensitive MCS search[21] is conducted with all potential reference ligands that have been cocrystallized with the protein of interest. An initial match is considered successful if the MCS region contains at least 7 non-hydrogen (heavy) atoms. All matched reference ligands are sorted by the MCS similarity metric. The MCS similarity of a compound pair is defined as the ratio of heavy atom count of the MCS region to the averaged heavy atom count of the molecule pair:

$$\mathrm{MCSratio} = \frac{2\mathrm{HA}_{\mathrm{mcs}}}{\mathrm{HA}_{\mathrm{template}} + \mathrm{HA}_{\mathrm{candidate}}} \times 100\%$$

$\mathrm{HA}_{\mathrm{template}}$, $\mathrm{HA}_{\mathrm{candidate}}$, and $\mathrm{HA}_{\mathrm{mcs}}$ are the heavy atom count of the reference template ligand, the candidate ligand, and the MCS region, respectively. The maximum possible value of MCS ratio is 100% if a candidate ligand and template ligand are identical. In the single docking mode ($N = 1$), only the top ranked template protein−ligand complex is selected. For ensemble docking, we retrieve the top N ranked ligand templates and the corresponding protein structures. In cases when no reference ligands are identified, the workflow can resort to a traditional, unbiased docking approach.

Once the template reference ligand(s) are identified, a conformational search is conducted to generate the conformer ensemble for the candidate ligand. The coordinates of the MCS region of the template ligand are directly transferred to the candidate ligand and restrained. Among all conformer generation engines we benchmarked, Omega[22] with the fixsmarts option (termed as Omega_MCS) allows for fast and accurate conformational space sampling while retaining the MCS region with exact coordinates as the X-ray template. Notably, Omega uses only one set of atom−atom mapping for

DOI: 10.1021/acs.jcim.5b00186
J. Chem. Inf. Model. 2015, 55, 1460−1468

a symmetrical MCS SMARTS pattern. This issue can potentially lead to an insufficient sampling of the conformational space for an asymmetrical candidate or reference molecules (Figure S1 illustrates the problem). We implemented an additional sampling step for Omega conformers to improve the sampling issue of molecules with symmetrical MCS regions. This step involves overlay of conformers onto a template ligand through all permutations of possible MCS atom−atom mappings and is followed by redundant conformer elimination.

Since the starting coordinates of the MCS region are already located inside the binding site, the second step, conformational generation, also serves as a first attempt for candidate ligand pose placement. Poses that strongly clash with the protein are filtered out. Clashes with crystallographic water molecules are treated differently. If a particular ligand conformer is within 2 Å to any water in the binding site, the water molecule is deleted and not included in energy calculations in the pose refinement stage. The remaining poses are further clustered using in-place RMSD of the non-MCS region. Representative poses from each cluster are retained and subjected to the final in-pocket refinement.

In-pocket refinement utilizes a local minimization of the final set of conformers in the protein pocket. During the minimization, atom-based flat-bottom potential restraints with a radius of 1.5 Å are imposed on the candidate MCS region. The default is to have the protein fixed while allowing active water molecules to rotate during the minimization. The user does have the option, however, to treat specific (i.e., user selected) residues to be flexible or restrained during minimization. After minimization, the interaction energy between each conformer of the minimized ligand and the protein is calculated using a docking score or MM-GBSA based scoring function. The top scoring pose is chosen as the final prediction. The protocol can, in principle, be used with any existing docking package like GLIDE or GOLD which can utilize positional restraints. In our current implementation, water and clash filtering, in-protein refinement is conducted with the MMFF94s[23] force field and a GBVI[24] solvent model using MOE.[25] A comparison of various forcefields, solvent models, and dielectric constants indicated minimal difference in terms of pose prediction accuracy (Figure S2). The final score is computed with MOE's GBVI/WSA[26] docking score. The entire process is executed in parallel for multiple ligand-protein templates if ensemble docking mode is chosen. In ensemble docking mode protein structure from the corresponding ligand template is chosen for pose refinement.

2.2. Validation Set Design. To estimate the accuracy of the protocol, we utilized a combination of public and internal X-ray complex data. The public data set contains 8784 candidate-reference ligand pairs covering 1742 ligands and 125 protein targets. The list was curated from the Protein Data Bank (PDB)[14] utilizing protocol described below (also see Figure S3). Only high resolution X-ray structures (≤2.5 Å) with at least one ligand bound (≤35 heavy atoms) were considered. No additional ligand filter was used. All proteins were prepared and minimized using a standard protein preparation protocol. In rare cases, this preparation failed and the structure was therefore removed from the set. For each protein target, all available protein−ligand X-ray structures were superposed to the same spatial reference using their backbone atoms. All associated ligands for a given protein target were then extracted and subjected to a pairwise MCS search. A ligand pair was defined if the MCS region contained at least seven heavy atoms. Ligand pairs likely binding to different regions, as indicated by their centroid distance greater than 8 Å after protein superposition, were filtered out. This filter accounted for less than 2.0% of the total reference pair entries. In order to produce statistically significant comparisons, we only considered proteins with more than 20 ligand pairs. This automated curation process could include a small number of poorly refined structures, as previously noted by PDB data quality survey.[27,28] However, the effect on the general conclusions or statistics should be minimal due to very large size of the validation set. Figure 2 shows the protein family statistics of the final set. A list of all pairs and PDB codes in this set is included in Table S1, while a per-target distribution of unique ligands is shown in Figure S4.

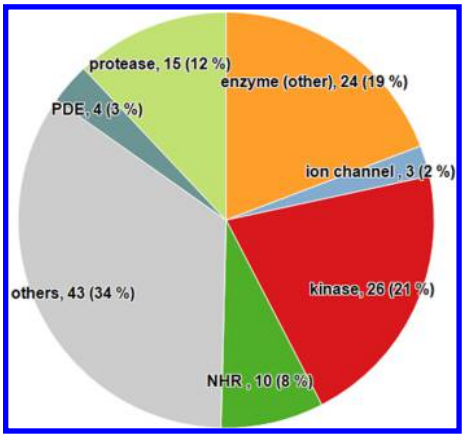


Figure 2. 125 proteins in the PDB validation set distributed by family. PDE: phosphodiesterase. NHR: nuclear hormone receptor.

The internal validation set, which was used for testing method coverage and chronological performance, consisted of seven small-molecule discovery efforts having a sufficient number of protein−ligand X-ray structures. The organization of the data was identical to the PDB set. The deposition date of each crystal structure and the earliest assay date of the compounds were also recorded, which allowed us to organize these sets chronologically. For both public and internal data sets, all compounds were extracted and converted to SMILES to avoid any potential conformation memory bias in the candidate ligand docking process.

3. RESULTS

3.1. Common Substructure Implies Similar Binding Mode. The first step in this process was to examine our premise that compounds sharing a common substructure will exhibit similar binding modes. The in-place RMSD of the ligand's MCS region was computed after protein backbone superposition. In more than 87% ligand pairs examined, the MCS RMSD is within 1.5 Å. This value increases to 89% if the MCS similarity ratio is at least 50%. The choice of width for flat-bottom MCS potential of 1.5 Å for the refinement stage was influenced by these statistics. Interestingly, as plotted in Figure 3, the average MCS RMSD is only weakly related to the MCS similarity of the ligand pair. Except in a limited number of cases, the ligand size itself does not have a strong impact on where the common substructure binds inside the protein (Figure S5). This observation is consistent with fragment-based approaches where the molecules are usually built from a smaller

DOI: 10.1021/acs.jcim.5b00186
J. Chem. Inf. Model. 2015, 55, 1460−1468

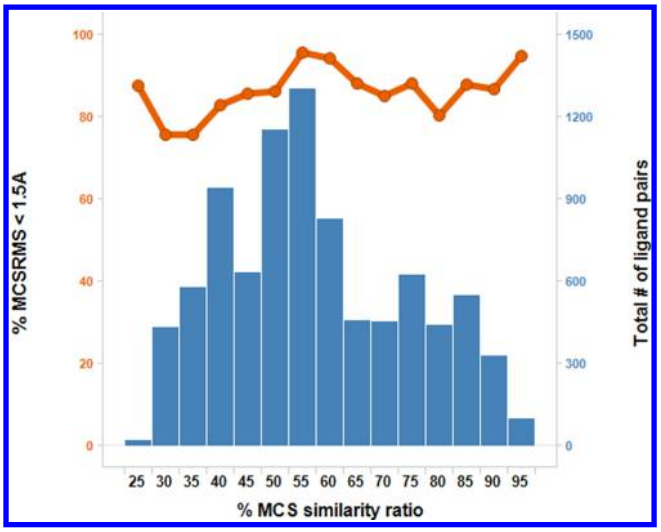


Figure 3. Relationship between MCS similarity ratio and the RMSD of the MCS region for each ligand pair in the data set. The blue bars show the number of ligand pairs in each MCS similarity ratio bin. The orange line shows the percentage of ligand pairs whose MCS regions are within 1.5 Å RMSD.

fragment core and by definition this core is a part of the MCS for the grown ligands.[18,29,30]

In 11% of the cases, the MCS RMSD of the two structurally related ligands exceeds 1.5 Å, hinting at the possibility of alternative binding modes, which was previously reported in the literature for certain targets with a fragment bound.[31,32] This data suggested that the upper limit of the docking success rate of any protocol relying on existing X-ray structures for this data set is approximately 89%. Indeed, alternative binding modes are one of the largest contributors to the docking failures and are discussed in more detail in the following section.

In addition to the MCS RMSD, we also computed a 3D ligand shape similarity between ligand pairs using ROCS.[33] As expected, compounds with a higher MCS ratio tend to be more similar in shape when they bound to the same protein (Figure S6). When the MCS region is ≥50% of the total molecule, in 77% of the cases two ligands shared a TanimotoCombo of greater than 1.0. This finding strongly suggests that two ligands sharing a large common substructure tend to have a similar shape inside the binding site and, hence, have a similar binding mode.

3.2. MCS Based Protocol Can Accurately Reproduce Ligand Bioactive Conformations. A key feature of a successful docking protocol lies in its ability to generate conformers that are similar to those found in co-crystal structures. This is particularly important since limited sampling is performed after the initial pose placement. To investigate the applicability of various conformer generators used in our protocol, we examined their ability to reproduce ligand bioactive conformations. Omega with MCS restraints (termed as Omega_MCS) was compared to the three non-MCS based methods. These included two knowledge-based protocols: the original Omega protocol (termed as Omega_noMCS), CCDC's conformer generator and one forcefield based method utilizing Macromodel.[34] The mixed torsional/low-mode sampling protocol along with the OPLS2.1 force field in Macromodel was used. The starting ligand conformation was regenerated from SMILES for each method. Different conformational generation methods may fail to produce conformers for some molecules. To ensure a valid comparison was performed with the identical data set, all failed cases were excluded. The lowest RMSD of any conformer ($RMSD_{min}$) to the bioactive conformer was examined for each method (Figure 4). Of the four methods, Omega_MCS is superior in retrieving

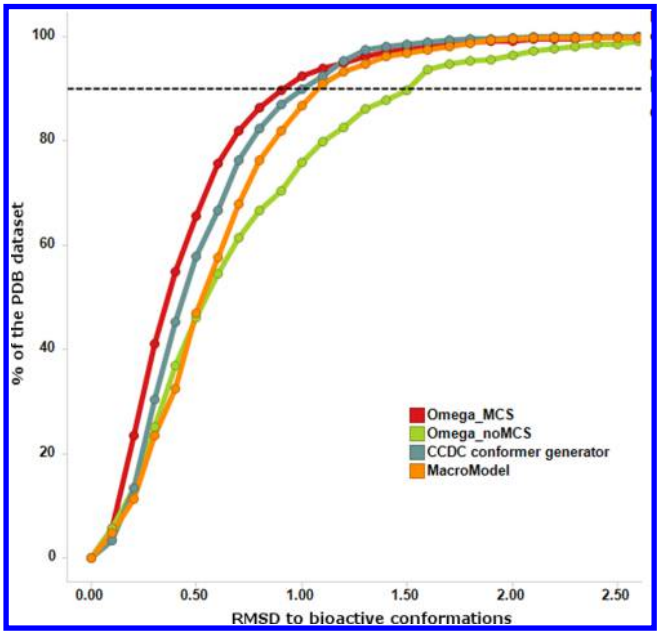


Figure 4. Ability of the conformer generators to reproduce bioactive conformations.

the bioactive conformer. In 90% of cases, at least one conformer within 0.90 Å from the bioactive conformation was generated. Note that the CCDC conformer generator also performed quite well (however, in the current version, incorporation of MCS restraints is not possible). In contrast to the other protocols, Omega_MCS has a smaller dependence on ligand flexibility as retraining the MCS region greatly reduced the space for conformational sampling (Figure S7). We conclude that the MCS based conformer generator can successfully generate conformer candidates close to the bound state conformation.

3.3. Docking Performance on PDB Validation Set. Following the conformer generation step, all conformers are subjected to clash filtering, conformer clustering, in-pocket refinement, and final scoring. The in-place RMSD is computed between the top scored pose and the X-ray ligand conformation and summarized in Table 1. For the commonly applied cutoff of 2 Å RMSD, the protocol has a 74.8% success rate for all 8784 cases tested. As shown in Figure 5, the docking accuracy is highly related to the candidate-template MCS similarity ratio emphasizing that accurate prediction can be anticipated when the candidate ligand and template ligand share a larger common substructure. Analysis of Figure 5 suggests the use of an MCS ratio ≥50% as a default requirement for selecting reference ligands from the library. Narrowing the analysis to this subset of 6198 pairs (Table 1) where the MCS ratio is ≥50% shows improved overall docking accuracy of 78.8%.

The statistics discussed above were based on all pairs of candidate-template ligands and probably represents a lower bound to most cases where MCS docking is used. In most lead optimization projects, multiple template ligands are often available. By leveraging these template ligands, we investigated

DOI: 10.1021/acs.jcim.5b00186
J. Chem. Inf. Model. 2015, 55, 1460−1468

Table 1. MCS Docking Performances Based on Different Template Selection Criteria[a]

MCS ratio	reference ligand selection	$N_{prediction}$	N_{ligand}	% < 1 Å	% < 2 Å
no cutoff	no selection, all pairs	8784	1742	48.1 ± 0.4%	74.8 ± 0.3%
≥50%	no selection, all pairs	6198	1331	52.2 ± 0.6%	78.8 ± 0.5%
≥50%	**most similar reference**	**2212**	**1331**	**58.2 ± 1.0%**	**81.5 ± 0.8%**
≥50%	ensemble docking	6198	1331	57.2 ± 1.3%	81.8 ± 1.1%
≥50%	most similar reference, medchem friendly ligands	1481	874	62.9 ± 1.2%	84.7 ± 0.9%

[a]Error bar is estimated from the standard deviation of docking accuracy in 1000 resampled dataset using bootstrapping.

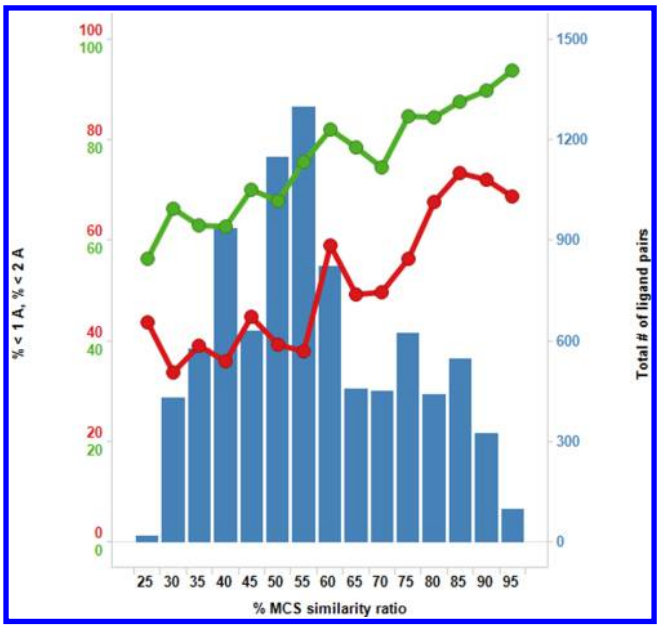


Figure 5. Plot of docking accuracy versus MCS similarity between candidate and reference ligands. The red and green lines show predictions within 1 and 2 Å, respectively. The blue bars show the number of ligand pairs in each MCS similarity ratio bin.

whether template selection heuristics or selecting the top scoring pose in ensemble docking can lead to improvements in accuracy. For the 1331 ligands (MCS ratio ≥50%), selection of the prediction from the template corresponding to the highest MCS ratio (thus narrowing down the original 6198 pairs to 2212 pairs) improves the docking accuracy to 81.5% (Table 1). Selection of the highest scoring pose from all ligand-protein templates (i.e., ensemble docking) gives a similar docking accuracy of 81.8%. This observation suggests that picking the reference protein−ligands with the highest MCS ratio is the method of choice for this protocol as it gives very similar accuracy to the more laborious ensemble docking. Finally, it is worth noting that certain ligands in the PDB set are not of high interest for drug development. Docking accuracy on 874 medicinal chemistry friendly ligands, i.e. those passing Lilly's medchem rules,[35] is 84.7% (Table 1, last row).

Docking accuracy results were further analyzed with respect to protein families (Table 2). Only the protein−ligand template with the highest MCS ratio to each candidate ligand was included in the analysis. In contrast to traditional docking protocols, the accuracy of MCS-based docking is not strongly dependent on protein families (Figure S8). The variability of docking accuracy among different target families can largely be explained by the differences in MCS similarity (Figure 5). Thus, the availability of highly similar templates appears to be the most critical factor for docking success. The exception, in this data set was the phosphodiesterase target family (Pde10a,

Table 2. MCS Docking Accuracy by Protein Families

family	$N_{prediction}$	% < 1 Å	% < 2 Å
ion channel	49	55%	76%
kinase	419	55%	83%
NHR	174	72%	93%
protease	128	48%	80%
PDE	39	38%	67%
other enzyme	370	52%	72%
other	1033	61%	83%

PDE4B, PDE4D2 and PDE5A). Analysis of this set of structures revealed that many of the failed cases arise from multiple symmetrical binding modes. The limited number of cases (39 pairs) from the PDB may also have influenced the statistics compared to the other families. Where more pairs exist, for example within Lilly's in-house test cases of two PDE projects, an accuracy of 90% or higher was achieved (See Figure S9).

The overall performance of this protocol is considerably better than standard commercial packages reported in the literature[11] for similar benchmark data set sizes. The accuracy of this protocol on kinases is comparable to the POSIT kinase benchmark set reported on OpenEye's Web site,[36] although with a different data set. A direct comparison was also carried out with one of the most widely used docking programs, GLIDE.[6] With an MCS cutoff of 50%, 6198 ligand pairs used in our protocol were subjected to GLIDE docking. To carry out the GLIDE docking, the same protein structures used in MCS docking were prepared with the Protein Preparation Wizard protocol and the GLIDE docking grids were generated using the bound ligand to define the size and position. In 796 (13%) cases, the GLIDE process failed to produce any ligand pose, including failures at the protein preparation, grid generation or the docking stages. These cases were excluded from the study. The results of the remaining 5402 cases are shown in Figure 6. For this reduced set, the MCS single template docking's accuracy of 80% (82% if most similar reference is used) is almost twice the accuracy of GLIDE (43%), demonstrating superiority of utilizing MCS restraints.

3.4. Analysis of Failed Cases. We inspected and categorized the 1315 (21%) cases where the MCS candidate-template similarity was at least 50%, but the protocol failed to produce docked structures within 2 Å. The majority of them resulted from the presence of an alternative binding mode, a scoring function issue or protein flexibility differences (Figure 7a). In the following section we provide further analysis of them and suggest improvements and/or prospective detection of potential failures.

3.4.1. Alternative Binding Mode. Further analysis reveals that 35% of the failed docking cases (7% of the total dockings) resulted from the common substructure for the candidate and template ligand exhibiting different binding modes (Figure 7b

DOI: 10.1021/acs.jcim.5b00186
J. Chem. Inf. Model. 2015, 55, 1460−1468

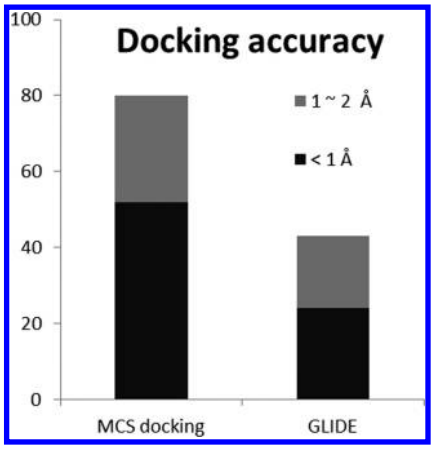


Figure 6. Overall docking accuracy for MCS docking and one non-MCS based docking protocol on the PDB validation set. Performance is computed for all structure pairs with an MCS cutoff of 50%. Detailed data is listed in Table S1.

exemplifies the alternative mode case). This ranged from a simple functional group rotation in a solvent exposed region of the binding pocket in a small number of examined cases to 180° rotation between the binding mode of the ligand and the template. The former can possibly be corrected by the removal of solvent exposed groups from the MCS region. Alternative binding mode(s) often existed in fragment-bound structures,[31] covalent bound structures and are frequently seen in certain kinases inhibitors. During the course of a drug discovery project, the prevalence of this phenomenon can be estimated from existing data and routinely monitored as the project progresses. If the prevalence is expected to be high, a decision can be made to use ensemble docking or to forego the MCS based approach.

3.4.2. Pose Ranking and Scoring Functions. In 46% of failed cases (9% of total cases), a native-like pose existed among all solutions but ranked lower than the top scoring pose. Among these, approximately 10% fall into a "close-miss" category, where the native-like solution was within 0.25 kcal/mol from the best scoring pose (the same cutoff of 0.25 kcal/mol was used in a previous publication[26]). In the other 36% of cases, the native-like solution ranked significantly lower in score (Figure 7c). Improving the latter cases will require an improved treatment of protein−ligand interactions and ligand strain energy, which is beyond the scope of this paper. Alternatively, ranking poses based on 2D/3D similarity to the template ligand, as used by POSIT, or some combination therein could also be considered.

3.4.3. Conformer Generation Failures. In 8% of the failed cases (2% of total examined cases), the Omega-MCS based conformational generator failed to produce a conformer close to the bioactive conformation for the candidate ligand. This was observed in cases where the template and candidate ligand shared a similar conformation with respect to its MCS region. In certain cases, the symmetrical matching issue (Supplemental Figure S1) was not fully solved by the additional sampling procedure (such as in certain PDE ligands). In other cases, sampling in the non-MCS region did not adequately explore

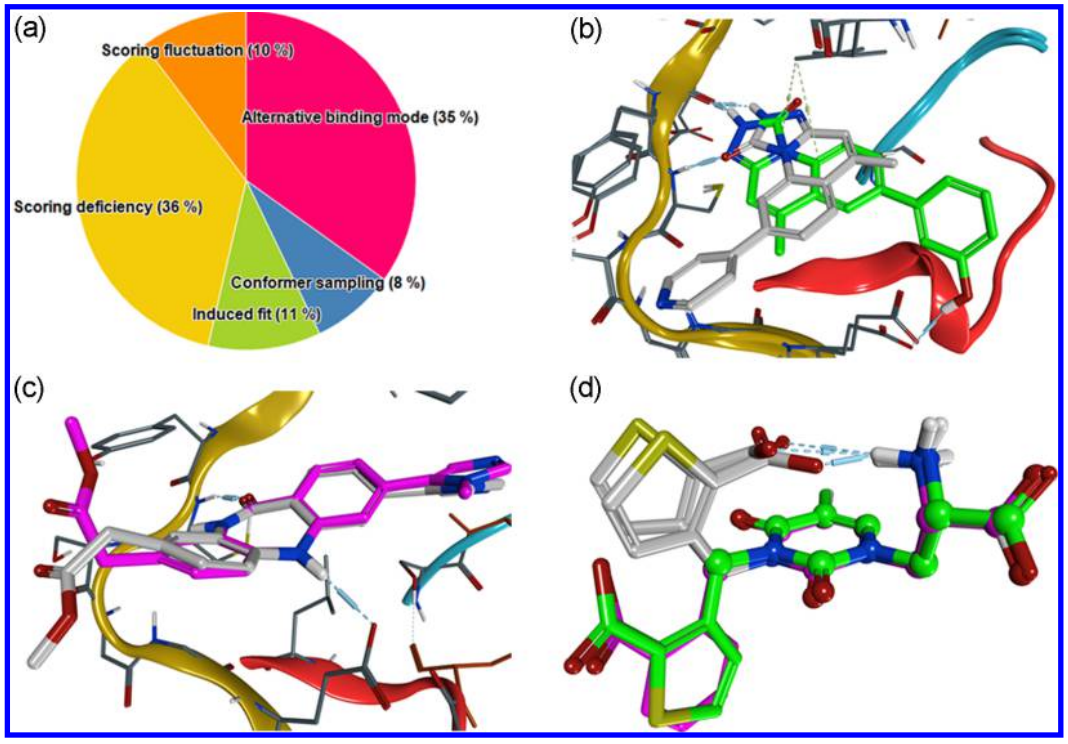


Figure 7. (a) Classification of all failed cases. (b) Example of alternate MCS arrangement. In two CHK1 kinases (PDB 2X8I, green carbon, and 2X8E, white carbon), the triazolone-containing tricyclic ring displayed two ways to interact with the hinge motif. (c) Example of scoring function deficiency. The CHK1 ligand from PDB 4FTT was predicted using protein 4FSY with its ligand as reference. The RMSD between the top predicted pose (purple carbon) and X-ray conformation (green carbon) is 2.2 Å, mainly due to the rotation at both ends of the molecule. Another docking solution (white carbon) ranked lower but was only 1.3 Å away from the correct solution. (d) An example of conformer generation failures. The clustered conformers (gray carbon) of the ligand in PDB 2F34 (green carbon) was predicted using the ligand in PDB 2H4G as template. The MCS region of the template ligand is shown in pink with ball and stick representation.

DOI: 10.1021/acs.jcim.5b00186
J. Chem. Inf. Model. 2015, 55, 1460−1468

the conformational space around the bioactive state (Figure 7d). As noted before,[37] an overestimated strain can potentially lead to insufficient sampling in these rare cases. These issues can be circumvented with additional sampling (e.g., through a variety of sampling methods available in Macromodel or MOE) given early detection of conformational sampling issues in a drug discovery project. For example, for each of the templates, the lack of conformational sampling around its own bioactive space is indicative of future sampling issues for candidate ligands.

3.4.4. Pose Placement Failure Due to Protein Induced Fit. The remaining 11% of failed predictions had native-like solutions in the conformer generation stage but either pose placement or final refinement moved the solutions outside the 2 Å native-like space. In most of these cases there was an induced fit effect, where the candidate protein had to adopt a different conformation from the original protein reference template to accommodate the ligand. This induced fit effect could range from side chain rearrangement to substantial rearrangement of protein backbone. We believe protein rearrangement is very difficult to account for and predict correctly during the docking process. Compounds which stabilize a unique protein rearrangement are often of high interest for structure-based efforts and can serve as good starting points for a unique pharmacological profile. Additional protein sampling is needed to identify alternative protein conformations that can fit the candidate ligand. While there are multiple ways to generate and evaluate protein conformational space, evaluating the accuracy of these sampling methods is beyond the scope of this work. In our hands, if reasonable poses cannot be obtained from a set of active compounds using MCS-based docking, these "not-fit-but-potent" compounds are often of the highest interest for the next batch of crystallography submission. Co-crystal structures are the ultimate resource to determine or verify alternative protein conformations generated by simulation. Additional reference protein−ligand templates can eventually help to cover the most relevant protein conformational space and further improve docking accuracy.

3.5. Applicability Domain of the MCS Docking Protocol in Medicinal Chemistry Programs. *3.5.1. Overall Protocol Coverage.* Since the protocol necessitates the presence of a common substructure between the candidate ligand and the reference set, it might not apply to all compounds for a given drug discovery project. To investigate the applicability domain of the protocol, we curated data from seven internal drug discovery projects: including three kinases, two PDEs, one NHR, and one protease project. We retrieved all co-crystal ligands and all compounds that have been tested in the primary assays for the selected targets. We calculated the fraction of active candidate compounds (<500 nM affinity) whose pose can be predicted with at least one X-ray structure as template. A prediction was defined to be possible if the candidate ligand shared more than 50% MCS similarity with at least one X-ray ligand. In all seven cases, the protocol covered more than 70% of candidate compounds across the tested data sets (Figure 8). Finally, the prospective docking accuracy was computed. Such accuracy analysis can only be performed on co-crystal structures therefore may introduce a sampling bias. Most of the compounds are designed assuming they share a similar binding mode to previous X-ray ligands. Less active compounds, which may have not fit well in the common modes, are usually not a subject for further X-ray studies. Nevertheless, the average docking accuracy of 82% is similar to

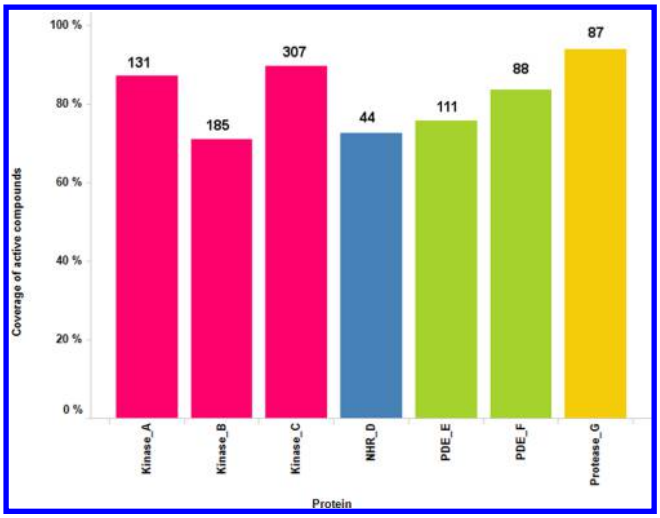


Figure 8. Coverage of active compounds (<500 nM in main in vitro assay). Coverage was computed as the percentage of compounds that can be predicted by the MCS docking protocol over all active compounds in the assay. The number above each bar reflects the total number of xray structures. See Table S2 for the raw data used to generate this figure.

the overall performance in the PDB data set, though varying by targets (Figure S9). In most cases, the model accuracy is relatively stable in the early phase of the project.

3.5.2. Chronological Coverage for Kinase A. The above analysis gives an upper bound of domain applicability for this set of targets. At the start of the project only few X-ray structures of complexes may be available. Therefore, the chronological coverage of the MCS docking protocol will likely change along the project time line. The chronological analysis allows us to better understand the applicability of the protocol to synthesized and designed compounds.

All project compounds, including the ones with available X-rays, were arranged chronologically. In a drug discovery project, binding pose predictions are applied on both synthesized and newly designed compounds. Prediction of compounds with known activity can augment the X-ray structures to better rationalize SAR. The prediction of newly designed compounds can help to evaluate the hypothesis and prioritize compounds for synthesis. Therefore, we examined two timeline metrics. "Retrospective overall coverage" represents the percentage of compounds whose poses can be predicted by X-ray templates available at that time the molecules are assayed. The second running metric, "prospective 3 month coverage" is defined as the percentage of candidate ligands that were assayed in the 3 months following the availability of each X-ray structure. This metric allows evaluation of the protocol's coverage of new design ideas.

Figure 9 shows the docking prediction coverage analysis for a kinase project A. For this project, the first 17 X-ray structures were available before synthesis of any compounds. These include apo structures and structures with ATP and its analogs as well as literature compounds. The retrospective coverage initially reaches 100% as analogs of literature compounds were synthesized. Between X-ray structure nos. 20 and 30, the protocol coverage dropped to 50% and then further down to 30%. This was due to the completion of a screening campaign which introduced greater chemical diversity. At this time many different scaffolds were explored simultaneously without the availability of structural information, and therefore the MCS

DOI: 10.1021/acs.jcim.5b00186
J. Chem. Inf. Model. 2015, 55, 1460−1468

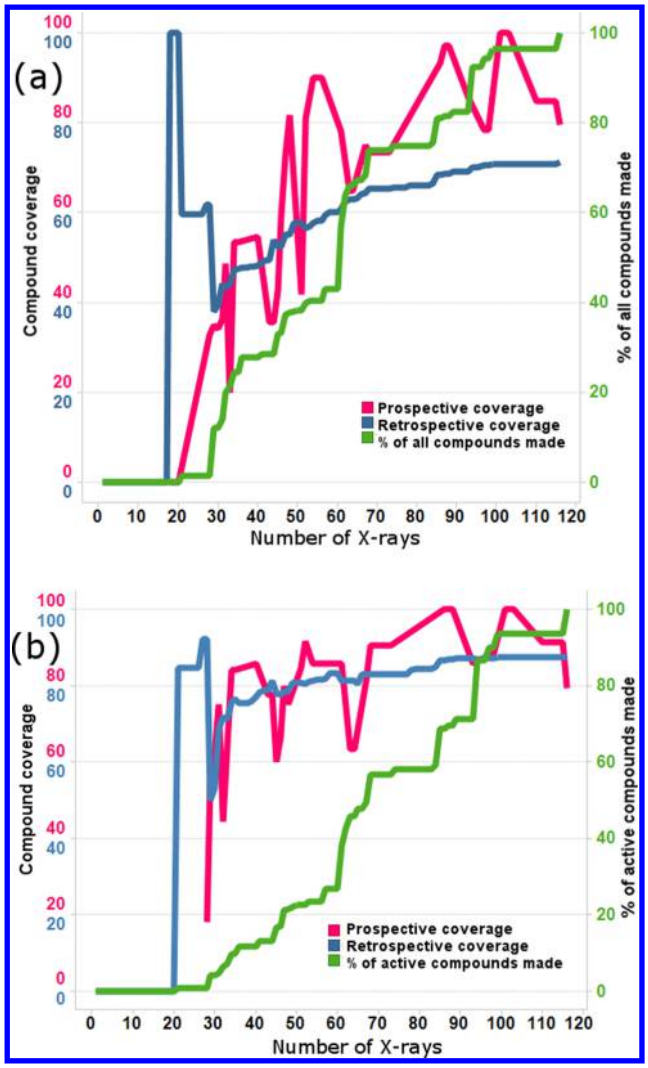


Figure 9. Methods to analyze docking prediction coverage for kinase project A. The *x*-axis is the number of available X-ray structures and is shown in chronological order, which is used as a surrogate to represent the project time line. The green line indicates the percentage of compounds made over the project time. The blue line shows the percent retrospective coverage and it correlated well with how this project evolved over time. In both figures, the pink curve represents prospective coverage and the blue curve represents retrospective coverage. The prospective 3 month coverage is computed for each chronologically ordered X-ray structure, and we assume all compounds synthesized and tested within the next 3 months need to be predicted using the first *N* X-ray structures. (a) Coverage on all tested compounds. (b) Coverage on active compounds (<500 nM in main in vitro assay) only.

based docking protocol had limited applicability. The main scaffold was selected around the time when structure no. 30 was determined (Figure 9a). This resulted in a gradual increase of the protocol's applicability in the project. After structure no. 55, the prospective coverage varied between 80% and 100% and retrospective coverage reached 65% and was maintained throughout the time of the project. The higher prospective coverage over the retrospective coverage arises from the fact that the latter part of the project was focused on optimization of compounds whose analogs had X-ray structures, while many earlier compounds, often less active, did not have crystallized analogs. The coverage for active molecules only (Figure 9b) is similar but higher than that for all compounds (Figure 9a). The largest increase in docking prediction coverage for active compounds also comes after structure no. 30 was solved. These X-ray templates represent a scaffold of high interest and were critical in designing the next generation analogs. As a result, both coverage metrics reached 75% and gradually plateaued as the project progressed. After structure No. 55, the retrospective coverage reached more than 80% of the active compounds.

4. CONCLUSIONS AND DISCUSSION

In this article we presented a validation and analysis of a maximum common substructure-based biased docking protocol. The protocol is applicable to any docking algorithm which can utilize restraints for specific chemical substructures. It was shown that an overall docking accuracy of 82% across tested cases was achieved and demonstrated an applicability domain for drug discovery projects varying from 70% to 90%. The docking success rate compares very favorably to one unbiased docking method, which was able to successfully dock 43% of the current data set. For specific drug discovery applications exemplified in this paper, both the success rate and the applicability domain make this algorithm an excellent choice for SAR rationalization and prospective ligand design. We have examined failed cases and found them to be largely driven by alternative binding modes, scoring limitations and induced fit effects. Lack of MCS coverage for new ligands or a consistently mispredicted binding pose can be used as one of the decision criteria for prioritizing ligands for crystallization. If no templates are available for a particular target or scaffold, the protocol can in principle be extended to utilize template ligands from other related targets, especially in target classes where binding pocket similarity is high. This work can also provide a foundation to other molecular docking[38] based applications including virtual screening and binding affinity estimations. Utilization of this protocol to systemically evaluate and improve binding affinity prediction methods is currently being investigated.

■ ASSOCIATED CONTENT

Supporting Information

Figures S1−S9 and Tables S1−S3. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00186.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: gao_cen@lilly.com.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank Joshua Guptill for help in organizing the X-ray database, Drs. Jeffrey Sutherland, Matthew Lee, Jon Erickson, and Jeremy Desaphy for useful comments and Dr. James Haigh (OpenEye) for helpful discussions regarding Omega.

■ ABBREVIATIONS

MCS, maximum common substructure; SAR, structure−activity relationship; RMSD, root mean square deviation; PDE, phosphodiesterase; NHR, nuclear hormone receptor

DOI: 10.1021/acs.jcim.5b00186
J. Chem. Inf. Model. 2015, 55, 1460−1468

REFERENCES

(1) Congreve, M.; Murray, C. W.; Blundell, T. L. Keynote review: Structural biology and drug discovery. *Drug Discovery Today* **2005**, *10*, 895−907.

(2) Huggins, D. J.; Sherman, W.; Tidor, B. Rational Approaches to Improving Selectivity in Drug Design. *J. Med. Chem.* **2012**, *55*, 1424−1444.

(3) Lybrand, T. P. Ligand—protein docking and rational drug design. *Curr. Opin. Struct. Biol.* **1995**, *5*, 224−228.

(4) 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*, 43−53.

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

(6) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177−6196.

(7) Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. DOCK 6: Combining techniques to model RNA−small molecule complexes. *RNA* **2009**, *15*, 1219−1230.

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

(9) Guedes, I.; de Magalhães, C.; Dardenne, L. Receptor−ligand molecular docking. *Biophys. Rev.* **2014**, *6*, 75−87.

(10) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in Molecular Recognition: The Effects of Ligand and Protein Flexibility on Molecular Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 45−55.

(11) Verdonk, M. L.; Giangreco, I.; Hall, R. J.; Korb, O.; Mortenson, P. N.; Murray, C. W. Docking Performance of Fragments and Druglike Compounds. *J. Med. Chem.* **2011**, *54*, 5422−5431.

(12) BIO-IT World http://www.bio-itworld.com/BioIT_Article.aspx?id=75442 (accessed Apr 6th, 2015).

(13) Hare, B. J.; Walters, W. P.; Caron, P. R.; Bemis, G. W. CORES: An Automated Method for Generating Three-Dimensional Models of Protein/Ligand Complexes. *J. Med. Chem.* **2004**, *47*, 4731−4740.

(14) Rose, P. W.; Prlić, A.; Bi, C.; Bluhm, W. F.; Christie, C. H.; Dutta, S.; Green, R. K.; Goodsell, D. S.; Westbrook, J. D.; Woo, J.; Young, J.; Zardecki, C.; Berman, H. M.; Bourne, P. E.; Burley, S. K. The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.* **2015**, *43*, D345−D356.

(15) Wu, G.; Vieth, M. SDOCKER: A Method Utilizing Existing X-ray Structures To Improve Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 3142−3148.

(16) Sutherland, J. J.; Nandigam, R. K.; Erickson, J. A.; Vieth, M. Lessons in Molecular Recognition. 2. Assessing and Improving Cross-Docking Accuracy. *J. Chem. Inf. Model.* **2007**, *47*, 2293−2302.

(17) Erickson, J. A.; Mader, M. M.; Watson, I. A.; Webster, Y. W.; Higgs, R. E.; Bell, M. A.; Vieth, M. Structure-guided expansion of kinase fragment libraries driven by support vector machine models. *Biochim. Biophys. Acta, Proteins Proteomics* **2010**, *1804*, 642−652.

(18) Erickson, J. Fragment-Based Design of Kinase Inhibitors: A Practical Guide. In *Fragment-Based Methods in Drug Discovery*; Klon, A. E., Ed., Springer: New York, 2015; Vol. *1289*, pp 157−183.

(19) *POSIT 3.1.0.5*; OpenEye Scientific Software, 2014; http://www.eyesopen.com (accessed Apr 6th 2015).

(20) Tuccinardi, T.; Botta, M.; Giordano, A.; Martinelli, A. Protein Kinases: Docking and Homology Modeling Reliability. *J. Chem. Inf. Model.* **2010**, *50*, 1432−1441.

(21) Raymond, J. W.; Willett, P. Similarity Searching in Databases of Flexible 3D Structures Using Smoothed Bounded Distance Matrices. *J. Chem. Inf. Model.* **2003**, *43*, 908−916.

(22) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572−584.

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

(24) Labute, P. The generalized Born/volume integral implicit solvent model: Estimation of the free energy of hydration using London dispersion instead of atomic surface area. *J. Comput. Chem.* **2008**, *29*, 1693−1698.

(25) *Molecular Operating Environment (MOE)*, 2013.08; Chemical Computing Group Inc., 2013; www.chemcomp.com (accessed Apr 6th 2015).

(26) Corbeil, C.; Williams, C.; Labute, P. Variability in docking success rates due to dataset preparation. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 775−786.

(27) Warren, G. L.; Do, T. D.; Kelley, B. P.; Nicholls, A.; Warren, S. D. Essential considerations for using protein−ligand structures in drug discovery. *Drug Discovery Today* **2012**, *17*, 1270−1281.

(28) Deller, M.; Rupp, B. Models of protein−ligand crystal structures: trust, but verify. *J. Comput.-Aided Mol. Des.* **2015**, DOI: 10.1007/s10822-015-9833-8.

(29) Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H. Fragment-Based Lead Discovery Using X-ray Crystallography. *J. Med. Chem.* **2005**, *48*, 403−413.

(30) Murray, C. W.; Verdonk, M. L.; Rees, D. C. Experiences in fragment-based drug discovery. *Trends Pharmacol. Sci.* **2012**, *33*, 224−232.

(31) Czodrowski, P.; Hölzemann, G.; Barnickel, G.; Greiner, H.; Musil, D. Selection of Fragments for Kinase Inhibitor Design: Decoration Is Key. *J. Med. Chem.* **2015**, *58*, 457−465.

(32) Babaoglu, K.; Shoichet, B. K. Deconstructing fragment-based inhibitor discovery. *Nat. Chem. Biol.* **2006**, *2*, 720−723.

(33) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, *50*, 74−82.

(34) *MacroModel, version 10.2*; Schrodinger LLC, 2014; www.schrodinger.com (accessed Apr 6th 2015).

(35) Bruns, R. F.; Watson, I. A. Rules for Identifying Potentially Reactive or Promiscuous Compounds. *J. Med. Chem.* **2012**, *55*, 9763−9772.

(36) OpenEye Document Page. http://docs.eyesopen.com/posit/theory.html (accessed Apr 6, 2015).

(37) Hawkins, P. C. D.; Nicholls, A. Conformer Generation with OMEGA: Learning from the Data Set and the Analysis of Failures. *J. Chem. Inf. Model.* **2012**, *52*, 2919−2936.

(38) Jain, A.; Nicholls, A. Recommendations for evaluation of computational methods. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 133−139.

DOI: 10.1021/acs.jcim.5b00186
J. Chem. Inf. Model. 2015, 55, 1460−1468